

管 理 科 学 基 础

吴育华 杜 纲 编著



A1012124

天 津 大 学 出 版 社

内 容 简 介

本书比较全面地介绍了管理科学的基本内容:规划技术、图与网络技术、决策技术、对策分析技术、随机运筹技术的概念、方法和模型,其中包括近年来的一些较新进展。书中内容深入浅出,每章后均附有习题,便于教学和自学。

本书可作为高等院校经济管理类和理工科各专业本科生、研究生的教材,也可作为经济管理及工程技术人员的自学参考书。

图书在版编目(C I P)数据

管理科学基础/吴育华,杜纲编著. —天津:天津大学出版社,2001.9(2002.10 重印)
ISBN 7-5618-1509-3

I. 管… II. ①吴…②杜… III. 管理学 IV. C93

中国版本图书馆 CIP 数据核字(2001)第 061253 号

出版发行 天津大学出版社
出 版 人 杨风和
地 址 天津市卫津路 92 号天津大学内(邮编:300072)
电 话 发行部:022-27403647 邮购部:022-27402742
印 刷 天津市宝坻县第二印刷厂
经 销 全国各地新华书店
开 本 787mm×1092mm 1/16
印 张 20.75
字 数 539 千
版 次 2001 年 9 月第 1 版
印 次 2002 年 10 月第 4 次
印 数 10 001—14 000
定 价 25.00 元

前 言

管理科学是一门应用科学、定量的方法去分析和解决管理决策问题的技术科学,其目的是帮助管理者在有限的资源条件下最优地实现组织目标,并为决策提供依据。随着管理现代化和科学化进程的不断加快,管理科学这一现代管理理论体系中的重要分支正在发挥着更大作用和得到更广泛的重视。由于这门学科不仅能够提供管理决策的科学方法,而且能够培养整体优化的决策思维方式,因此,我国各高等院校的经济管理类专业均普遍把这门学科作为主干课程。

近年来,我国高等管理教育有了相当大的发展。这种发展不仅体现在纵向的多层次性(本科、硕士、博士等)和横向的多专业性(管理科学与工程、工商管理、技术经济等),而且体现在对管理类专业的社会需求(在职人员申请硕士学位、管理人员培训、非管理专业学生的管理选修课等)的大量增加。因此,编写一本适应面较宽的管理科学教材就成为十分现实的需求。本书就是为适应这种需求编写的。

本书比较全面地介绍了管理科学的基本内容,其中包括近年来的一些较新进展。书中内容深入浅出,便于教学和自学,同时强化了绪论的地位,有利于学生对这门学科及其在管理决策问题中的应用程序等的整体了解。

本书作者都是多年从事这门学科教学的教师。但由于时间比较仓促,书中难免出现一些不当之处,敬请读者批评指正。

在本书的编写和出版过程中,参考了一些国内外的相关文献,得到了天津大学教务处和天津大学管理学院的大力支持。管理学院的范贻昌教授对本书内容提出了宝贵意见,博士生曾祥云、刘喜华、魏隼、安利平、王初等同志做了大量工作,谨在此一并表示衷心的感谢。

作 者

2001年1月

目 录

I	引论	(1)
	第1章 绪论	(1)
	1.1 管理与管理科学	(1)
	1.2 管理科学的工作程序	(4)
	1.3 管理科学的主要内容及本书结构	(7)
II	规划技术	(10)
	第2章 线性规划	(10)
	2.1 线性规划的模型与图解法	(10)
	2.2 单纯形法	(16)
	2.3 对偶问题与灵敏度分析	(30)
	2.4 运输问题	(40)
	2.5 线性整数规划	(49)
	第3章 非线性规划	(64)
	3.1 基本概念	(64)
	3.2 无约束极值问题	(69)
	3.3 约束极值问题	(77)
	第4章 多目标规划	(87)
	4.1 多目标规划模型及其解的概念	(87)
	4.2 多目标规划的解法	(91)
III	图与网络技术	(111)
	第5章 图与网络分析	(111)
	5.1 基本概念	(111)
	5.2 最小支撑树问题	(113)
	5.3 最短路问题	(115)
	5.4 最大流问题	(119)
	第6章 网络计划	(129)
	6.1 网络计划图的绘制	(129)
	6.2 时间参数计算与关键路线确定	(134)
	6.3 网络图的调整及优化	(139)
IV	决策技术	(150)
	第7章 风险型决策	(150)
	7.1 决策问题的基本要素及分类	(150)
	7.2 风险型决策	(152)
	7.3 完全不确定情况下的决策	(172)
	第8章 库存决策	(178)
	8.1 库存决策的基本概念	(178)
	8.2 确定型库存模型	(179)

8.3 随机型库存模型	(186)
第9章 多阶段决策——动态规划	(197)
9.1 多阶段决策问题	(197)
9.2 动态规划的基本概念	(198)
9.3 最优性原理与递推方程	(200)
9.4 动态规划应用例解	(202)
第10章 多目标决策	(221)
10.1 特尔非法	(221)
10.2 层次分析法	(222)
10.3 数据包络分析法	(229)
V 对策分析技术	(241)
第11章 二人有限零和对策	(241)
11.1 基本概念	(241)
11.2 二人有限零和对策的纯策略对策模型	(243)
11.3 二人有限零和对策的混合策略对策模型	(246)
第12章 二人有限非零和对策	(253)
12.1 非合作的二人有限非零和对策	(253)
12.2 合作的二人有限非零和对策	(258)
VI 随机运筹技术	(262)
第13章 排队系统分析	(262)
13.1 基本概念	(262)
13.2 到达与服务的规律	(265)
13.3 M/M/1 排队模型	(268)
13.4 M/M/C 排队模型	(277)
13.5 一般服务时间的 M/G/1 排队模型	(282)
13.6 排队系统的优化	(284)
第14章 马尔可夫分析	(289)
14.1 引言	(289)
14.2 关于正规随机矩阵的一些知识	(291)
14.3 马尔可夫链	(293)
14.4 吸收马尔可夫链	(300)
第15章 随机模拟技术	(310)
15.1 模拟过程的主要步骤	(310)
15.2 一个人工模拟的例子	(312)
15.3 随机变量模拟数据的产生方法	(315)
15.4 模拟语言简介	(319)

I 引 论

第 1 章 绪 论

1.1 管理与管理科学

1.1.1 管理与决策

随着现代科学技术和社会经济的不断发展,各个领域对管理的重视程度及对管理科学化的要求越来越高,各种组织(如企业)面对日趋复杂的生存环境,也必须更加精心地处理各种管理问题。

那么,什么是管理?管理就是管理者运用各种资源达成某既定目标的过程。在这一过程中,管理者为了更有效地运用有限的资源以更高水平达到目标,必须不断地做出各种决策。可以说,管理的过程也就是不断地进行各种决策的过程。

尽管决策的正确性不仅依靠科学而且凭借经验与艺术,但随着决策的难度以及决策失误后造成的损失程度的不断增大,那种仅凭经验与艺术的决策情形越来越少,即使是以往认为主要靠经验和艺术的那些非程序化或高层次决策,也往往要先经过一系列基于科学方法的信息处理和可行性研究。管理科学正是为管理决策提供科学方法的一门学科。随着信息化时代的到来,它必将在管理科学化的进程中发挥出更大的作用。

1.1.2 管理科学及其发展简史

1. 管理科学及其在科学技术体系中的地位

关于管理科学,可以有广义和狭义两种理解。广义的理解,认为管理科学(Management Sciences)是一门应用多学科与多领域理论、方法、技术和知识的综合性交叉学科,其目的是研究人类利用有限资源实现组织目标的管理活动方面的动态、复杂和创新的社会行为及其规律。按照广义的理解,管理科学实际上可以包含多门学科,其中既有以定性分析为主的组织行为学和企业战略管理,又有以定量分析为主的运筹学和计量经济学。而狭义的理解,则主要涉及广义范畴中的运筹学等定量部分,认为管理科学(Management Science)是一门应用科学、定量的方法去分析和解决管理决策问题的技术科学,其目的是帮助管理者在有限的资源条件下最优地实现组织目标,并为决策提供依据。它在现代科学技术体系的四个层次中属于技术科学层次,和它相应的基础科学主要包括数学、管理学和经济学等,作为它的哲学指导观的主要是唯物论和系统观,而和它相应的工程技术主要是管理工程等,如图 1.1 所示。

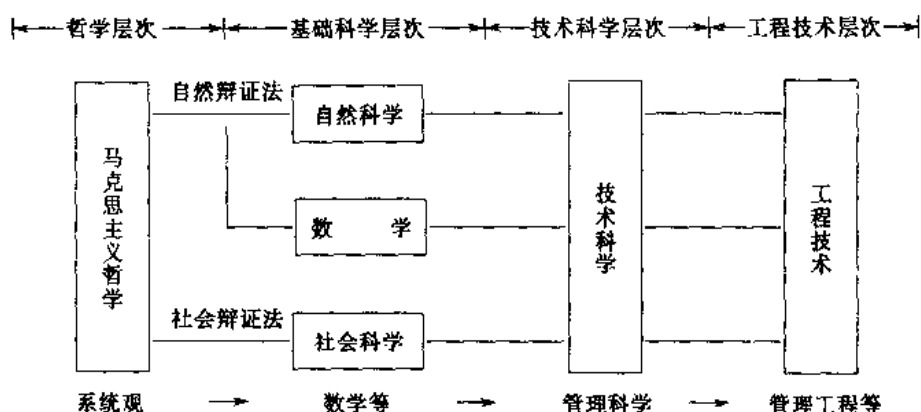


图1.1

本书是基于狭义的管理科学定义,定名为《管理科学基础》。这里面有两层含义:一是从狭义管理科学的角度,它包括了管理科学的基础内容;二是从广义管理科学的角度,它是基于系统观的整体优化思想和基于问题导向的科学方法论,以及基于数学模型的定量分析工具等一整套内容体系构成整个广义管理科学的重要基础。

2. 管理科学的发展简史

管理科学的萌芽,可以追溯到19世纪末至20世纪初。这一时期代表性的工作有二:一是泰勒(Taylor)提出了著名的科学管理理论;二是产生了若干将数学模型应用于管理的成果,如爱尔朗(Erlang)提出的排队模型和哈里斯(Harris)提出的EOQ存贮模型等。

作为一门独立的学科,管理科学产生于20世纪40年代。首先是在40年代初期的二次世界大战中,英国为了解决雷达站同整个防空作战系统的协调配合问题,成立了由多学科科学家组成的研究工作小组。工作小组卓有成效的工作使得在英军每个指挥部都成立了这样的小组。随后在美国军队中也建立了类似的组织。这些小组在确定扩建舰队规模、开展反潜艇战的侦察和组织有效的对敌轰炸等方面做了大量研究,为取得反法西斯战争的胜利及管理科学各有关分支的建立作出了贡献。对于这些小组的研究工作,英国称之为operational research,美国称之为operations research,即我国后来译成的“运筹学”。

二次大战以后,运筹学的研究和应用由军方扩展到民间,这门学科的理论体系也不断完善。1950年,英国伯明翰大学正式开设了运筹学课程,同年,第一本运筹学杂志《运筹学季刊》(O.R. Quarterly)于英国创刊。1951年,美国的莫尔斯(P. M. Morse)和金博尔(G. E. Kimball)合著的《运筹学方法》一书正式出版。1952年,美国运筹学会成立,并于同年出版了运筹学杂志(Journal of ORSA)。所有这些,标志着这门学科基本形成。

50年代以后,随着电子计算机技术的迅速发展,运筹学中的很多方法,如线性规划、动态规划等由于能够更加方便地求解而被进一步应用于实际管理系统的优化问题中。至50年代末,美国已有约半数的大公司在自己的经营管理中应用了运筹学,主要用于生产计划、物资储备、资源分配、设备更新等方面的管理决策。运筹学这门学科,也因其与管理决策的密切联系,以及在解决管理决策实际问题中逐渐形成了一整套系统、定量的科学方法,而被广泛称之为“管理科学”。管理科学的一些主要分支的产生年代如图1.2。

管理科学经过半个多世纪的发展历程,不仅在理论研究上具有了相当的深度和广度,而且其方法与信息技术相结合,形成了各种商业应用软件,使其具有更加广泛的应用。我国从50

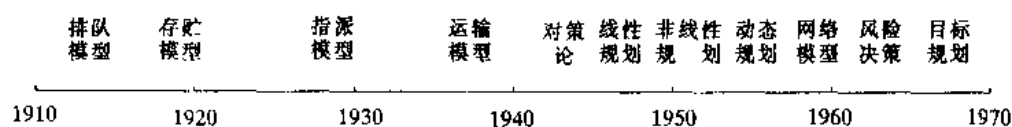


图1.2

年代中期开始研究这门学科以来,也取得了许多令世人瞩目的研究成果。全国各高等院校的经济管理类专业均普遍把这门学科作为主干课程列入教学计划,工商管理硕士(MBA)的教学计划中也把这门学科定为核心课程。我国还于1999年8月在北京成功地举办了第15届国际运筹学会联合会学术大会。相信这门学科必将在21世纪的知识经济与信息时代中进一步展现其生命力并不断得到发展。

1.1.3 管理科学的基本特性

管理科学具有以下的一些基本特性。

1. 以管理决策为基点

管理科学是一门为管理决策提供科学方法的学科。管理决策中的实际问题,既是管理科学产生的源泉,又是管理科学的应用对象。表1.1例举出管理科学方法直接应用于企业管理决策问题的例子。

表 1.1

管理科学方法	应用于企业管理决策的典型例子
线性规划	生产结构优化
非线性规划	投资组合优化
0-1 规划	选址问题
动态规划	资源分配问题
网络计划	工程计划优化
排队论	服务系统优化
存贮论	订货库存管理
风险型决策	机会选择

2. 以科学方法论为依据

管理科学是一门将“科学”方法论应用于管理决策的学科。所谓科学方法论是指进行科学研究的一般程序与方法,一般步骤如图1.3。

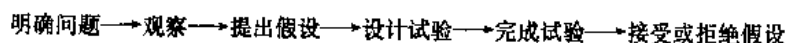


图1.3

应用管理科学方法解决管理决策中实际问题的一般程序也可分为类似的6个步骤,即明确问题、将问题归类、建立数学模型、求解模型、结果分析与模型检验、实施。这将在本章下一节详细阐述。

3. 以系统观点为指导

管理科学是以系统观为指导的一门技术科学。系统观对管理科学有两方面指导作用。一方面是用系统的观点看问题,将所研究的事物看做一个系统,看做一个具有特定目标、功能、结

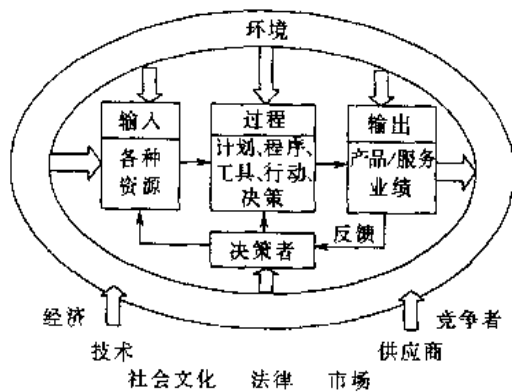


图 1.4

构和环境的系统,且系统与环境之间还应有输入和输出。图 1.4 给出了一个企业系统的例子。另一方面是对系统的“整体优化”。如果把管理科学多分支的丰富内容体系比作一棵大树,那么大树的主干就是“最优化”,大树的分支就是最优化在各个角度上的表现,大树的树根扎根于基础科学的土壤,而大树的果实则表示在各管理决策领域中的应用成果。可以说,整个管理科学方法的核心思想就是“整体优化”,而这正是系统观的精髓之一。

4. 以数学模型为主要工具

管理科学是一门以定量分析为主、以各种数学模型为主要工具的学科。数学模型是现实系统的一种简明映象,以数学模型为主要工具的定量分析方式不仅能够使人们对系统的认识更加深刻,而且便于使用计算机,从而能够节约费用和时间。管理科学的每一个分支都有自己特定的数学模型和基于模型计算的分析方法。因此,学习管理科学方法以及学会运用管理科学方法去解决实际问题的一个关键是学会建立各种数学模型。这将在本章下一节中阐述。

1.2 管理科学的工作程序

应用管理科学方法去解决管理决策中实际问题的一般工作程序可分为 6 个步骤,如图 1.5 所示。

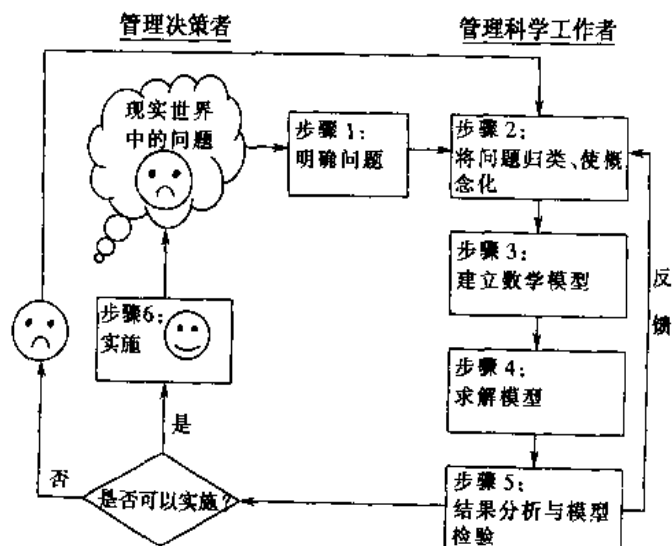


图 1.5

1.2.1 步骤 1: 明确问题

运用管理科学方法去解决管理决策实际问题的首要步骤是明确“问题是什么”,从而正确

地定义问题。这就需要与管理决策者进行沟通,对他提供的问题状况进行认真研究和系统分析。要明确究竟解决什么事情?为此必须做出什么决策?最终要达到什么目标?有哪些不可控的重要影响因素等等。

1.2.2 步骤2:将问题归类、使概念化

问题明确以后,还要对其做类型判别,即先看它是否属于能用管理科学方法解决的问题。如果是,再看它是属于哪一类管理科学问题,或者说它适合于用哪一类管理科学方法解决。下面例举了一些典型的管理科学问题类型。

1)分配问题 这类问题产生于:①存在若干要完成的任务;②存在两个以上完成这些任务的方案;③为完成任务所需的资源是稀缺的。问题是寻求资源的最佳利用方案,以使总效益最大。

2)竞争问题 这类问题产生于:决策行动的结局不仅依赖于决策者自己,而且依赖于竞争对手的行动,每个竞争者都想取胜。问题是寻求使每个竞争者都能接受的总损失最小的行动方案。

3)运输与指派问题 运输问题产生于:将某种物品(如矿产品或粮食)由若干产地运往若干需求地,问题是寻求使总运输成本最小的调运方案。还有一些问题(如生产安排问题)也可归为运输问题。指派问题是一种特殊的运输问题。它产生于:将若干项任务分派给若干人,每项任务由一个人完成,每个人也只能完成一项任务,问题是寻求使总效率最高的分派方案。

4)库存控制问题 原材料、资金、成品、人才等的适当储存或储备水平,往往是许多组织中的一个主要问题。尽管储存需要付出成本,但却可以从防止缺货损失及大批量购买的价格折扣中获益。库存控制问题是确定适当的订货间隔和订货数量,以使总的费用最低。

5)网络与工程计划问题 网络问题的种类很多,如寻求网络中两点的最短路和合理安排网络上的流量使总的通过量最大。复杂工程项目的计划与控制也可化为网络问题,求得工程的最小工期和关键工序。

6)排队与服务系统优化问题 当顾客为寻求某种服务而到达服务系统时,往往会产生排队现象,服务系统为减少排队现象而提高服务速率或增加系统容量需付出各种成本,一旦系统空闲又将受到损失。问题是如何合理地设置系统的服务速率及容量,以兼顾系统与顾客双方的利益。

7)预测问题 这是管理中经常遇到的问题,即希望知道在不同条件下系统将产生怎样的行为状态。

每一类管理科学问题中都包含着十分丰富的内容或可再划分为各种类型。根据问题的类型,可选择适当的管理科学方法来解决。当方法选定后,还要将问题概念化,即用所选定方法中的术语重新描述问题,或者说将方法中的要素与问题对号,为建立数学模型做准备。

1.2.3 步骤3:建立数学模型

建立数学模型是管理科学方法的关键步骤,具体内容主要是在问题概念化的基础上进一步确定模型的构成要素以及它们之间的联系,用变量和数学关系式表达出来,形成数学模型。

1.模型的构成要素

一般来说,模型的最基本构成要素由三部分组成,即结果变量、决策变量和不可控变量。

结果变量反映了系统达成目标的有效性程度,它依赖于决策变量的取值,是一种相依变量。决策变量描述了决策问题中必须作出选择的要素,是一种可控的独立变量。不可控变量是指系统环境中对决策有重要影响但不可控的因素,它也是独立变量。表 1.2 给出这三种变量的实际例子。

表 1.2

领 域	决策变量	结果变量	不可控变量
金融投资	投资数量 投资期限 投资时机	总利润 收益率 每股净利	通货膨胀率 贷款利率 竞争
市 场	广告预算	市场份额	可支配收入
生 产	产量 库存水平	总收入 质量水平	市场价格 技术
财 会	审计计划 计算机使用	数据处理成本 错误率	法规要求 计算机技术
运 输	装运量	运输总成本	运送距离规定
服 务	服务员数量	顾客的满意度	服务需求

2. 模型的结构

模型的基本构成要素由数学关系式联系在一起,便形成数学模型。模型的一般结构如图 1.6 所示。

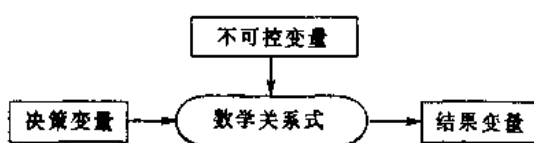


图1.6

管理科学的分支不同,模型的具体表现形式也不尽相同,其中最具有代表性的是数学规划模型。下面例举线性规划的例子说明数学模型的具体结构。

线性规划模型中的数学关系式有两种类型,即目标函数与约束条件。以一个生产结

构优化问题为例,它的目标函数表达式为

$$R = p_1 x_1 + p_2 x_2$$

式中, R 表示生产总收入; x_1 和 x_2 表示两种产品的产量; p_1 、 p_2 表示两种产品的价格。目标是极大化总收入。为达成目标要受到来自各方面的约束,如市场需求量的约束,表示为

$$x_1 + x_2 \leq 50$$

这里,50 是两种产品的总需求量上限。此模型的结构如图 1.7 所示。

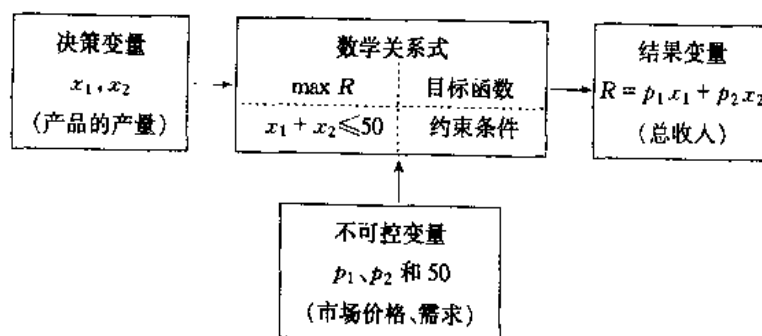


图1.7

1.2.4 步骤4:求解模型

管理科学的模型种类很多,解的性质及其求解方法各异。“最优解”是解的主流类型,它是在模型的所有可行解中找出最优的一个。除了“最优解”之外,还有其他一些解的类型,如“满意解”和“描述性的解”。“满意解”是指在模型的次优解或“各有所长”的一组解中使决策者满意的解;“描述性的解”相应于描述性的模型,该类模型的目的主要是描述系统在不同条件下的状态,可用于预测和分析系统的行为特征。相对于“最优解”,这些其他类型的解不妨称为“非最优的解”。

求解方法可分为“数值的”和“分析的”两种类型。数值方法一般是通过用某种模式一步一步搜寻并不断改进解的过程来求解,而分析方法是由数学公式一步求出解。图 1.8 给出管理科学主要分支模型的解及其解法的例子。

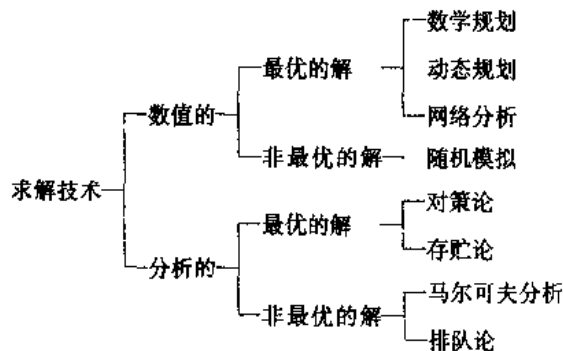


图1.8

1.2.5 步骤5:结果分析与模型检验

求得模型解以后,还要对它进行分析,同时要检验模型的正确性。分析和检验的内容首先是看结果与实际情况是否相符,模型是否能够正确地反映实际问题,具体方法是可将不同条件下的数据(如不同时期的数据)代入模型看相应的解是否符合实际。其次要分析模型中的参数发生小范围变化时对解的影响。这一分析称做灵敏度分析。如果解对参数变化的反应不过于灵敏,则在实际中可能会更容易应用。

经过结果分析与模型检验,如果认为模型不能很好地反映实际问题,则要重新对问题进行分析并适当修改模型。

1.2.6 步骤6:实施

对模型及其解分析和检验后,还要对其实际意义给以解释并将其提交给管理决策者。如果管理决策者认为可以实施,则管理科学的分析结果得以实现;如果认为仍不能实施,则要再次对问题进行分析并适当修改模型。

1.3 管理科学的主要内容及本书结构

1.3.1 管理科学的主要内容

管理科学是一门具有多分支内容的极其丰富的学科,主要内容可以归纳成下面的几大部

分。

1. 规划技术

规划技术主要是指数学规划技术,包括线性规划、非线性规划、整数规划和多目标规划等,是管理科学中最有代表性的主体内容。数学规划是在一定约束条件下寻求使目标极大化或极小化的最优解或满意解,其实际背景是在有限的资源条件下通过合理计划而实现总效益的最大化问题。

2. 图与网络技术

图与网络技术包括图与网络分析以及网络计划,是研究图与网络问题的分析与优化的管理科学分支。主要方法是基于图的基本理论对网络问题进行分析,求得使网络效益最大(如最短路、最大流、最短的项目工期、成本最低的项目工期)的最优解。

3. 决策技术

决策技术包括风险型决策、库存决策、多阶段决策和多目标决策等。其中风险型决策是研究在不确定条件下进行方案选择问题的管理科学分支,主要方法是基于收益或效用以及概率的基本理论,对方案的平均效益进行比较,求出最优方案。库存决策是研究库存计划与订货策略的管理科学分支,主要方法是建立各种类型的存贮模型,求得使总成本最少的最优解。多阶段决策又称动态规划,是研究多阶段决策问题的管理科学分支,主要方法是通过将问题划分阶段并逐段递推求解,最终求得使总效益最大的各阶段最优解。多目标决策是研究多准则评价与决策问题的管理科学分支,从各种不同的角度,采用定性与定量相结合的分析方法,对系统进行综合评价或排序。

4. 对策分析技术

对策分析技术是研究在对抗与竞争条件下的决策问题的管理科学分支,主要方法是建立各种类型的对策模型,求得使双方或多方达到平衡的最优解。其基本类型是二人有限零和对策及二人有限非零和对策。

5. 随机运筹技术

随机运筹技术包括排队系统分析、马尔可夫分析和随机模拟技术等。排队系统分析又称排队论,是研究随机服务系统的性能、状态及优化问题的管理科学分支,主要方法是建立各种类型的排队模型,求得在各种条件下反映系统性态的描述性的解。马尔可夫分析是研究由随机变量现时的运动状况来分析预测该变量未来运动状况的管理科学分支,主要方法是基于概率和随机过程的理论,通过系统状态和转移规律求得未来的状态。随机模拟技术又称系统仿真,是研究对静态离散的随机系统进行模拟分析的管理科学分支,主要方法是通过随机数和系统的有关概率分布对系统进行状态模拟。

1.3.2 本书的结构

本书由引论、规划技术、图与网络技术、决策技术、对策分析技术和随机运筹技术六部分组成,共 15 章,结构如图 1.9 所示。

从各部分内容的知识衔接关系来看,第一部分引论和第二部分规划技术是全书的基础内容,其他各部分内容大部分是并列的。图 1.10 给出本书的阅读顺序,其中的数字表示章的序号。

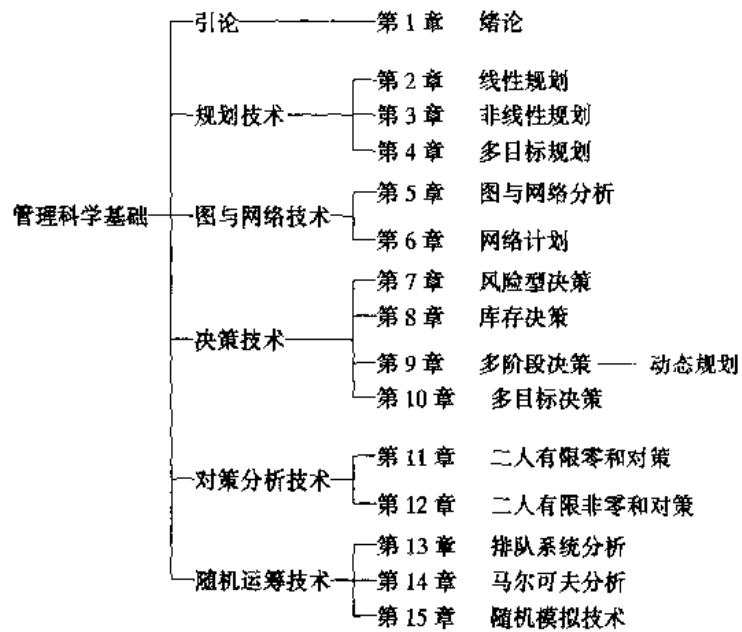


图1.9

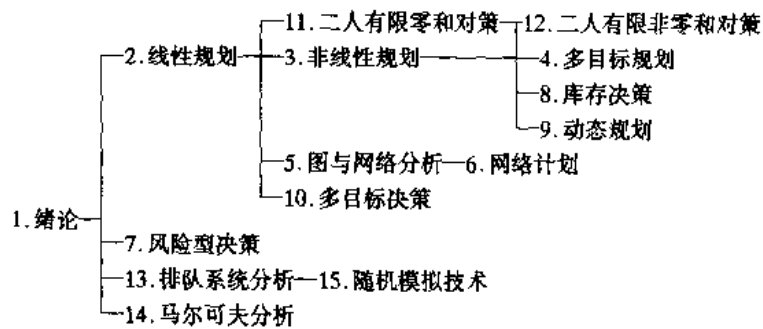


图1.10

II 规划技术

第2章 线性规划

2.1 线性规划的模型与图解法

2.1.1 问题的引入

要想在工农业生产、交通运输、商业贸易等各方面提高效益,有两种途径:一是革新技术;二是改进生产组织与计划,即合理安排有限的人力和物力资源,最合理地组织生产过程。数学规划能够为更好地配置资源、组织生产提供理论和方法,包括线性规划、非线性规划、整数规划、多目标规划等很多分支。其中线性规划是在现代管理中应用最广、理论比较完善的一个部分。随着电子计算机的发展,数学规划在现代化管理中的重要性日益明显。

问题1 某工厂计划生产甲、乙两种产品,生产1 kg 甲产品需耗煤9 t、电力4 kW·h、油3 t;生产1 kg 乙产品需耗煤4 t、电力5 kW·h、油10 t。该工厂现有煤360 t、电力200 kW·h、油300 t。已知甲产品每千克售价为7万元,乙产品每千克售价为12万元。在上述条件下决定生产方案,使总收入最大,具体数据如表2.1所示。

表 2.1

每 千 克 耗 量 资 源	产 品		资源限量
	甲	乙	
煤(t)	9	4	360
电(kW·h)	4	5	200
油(t)	3	10	300
单位价格(万元)	7	12	

为求解上述问题,设 x_1 为甲产品生产量, x_2 为乙产品生产量,称 x_1 和 x_2 为决策变量,向量 $[x_1 \ x_2]^T$ 表示一个生产方案,称之为决策变量向量,简称决策变量。由于煤、电、油资源的限制,生产方案 $[x_1 \ x_2]^T$ 的取值也必受到一定限制。例如,采取生产方案 $[x_1 \ x_2]^T$,耗煤总量为 $9x_1 + 4x_2$,它不能超过煤资源总量。这样所取的生产方案 $[x_1 \ x_2]^T$ 必须满足条件

$$9x_1 + 4x_2 \leq 360 \quad (\text{煤资源限制})$$

同样, $[x_1 \ x_2]^T$ 必须满足条件

$$4x_1 + 5x_2 \leq 200 \quad (\text{电资源限制})$$

$$3x_1 + 10x_2 \leq 300 \quad (\text{油资源限制})$$

同时, x_1 和 x_2 不能取负数, 即须满足

$$x_1 \geq 0, \quad x_2 \geq 0 \quad (\text{非负限制})$$

所以, 一个可行生产方案 $[x_1 \ x_2]^T$ 必须满足条件

$$\begin{cases} 9x_1 + 4x_2 \leq 360 \\ 4x_1 + 5x_2 \leq 200 \\ 3x_1 + 10x_2 \leq 300 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \quad (2.1)$$

式(2.1)称为一组约束条件。

对于每一个可行生产方案 $[x_1 \ x_2]^T$, 均有一个由这个方案产生的总收入 z 与之对应, 即有

$$[x_1 \ x_2]^T \rightarrow z = 7x_1 + 12x_2$$

称 $z = 7x_1 + 12x_2$ 为目标函数, 并由它衡量各生产方案的优劣。

综上所述, 对问题 1 求解最优方案的问题可由下列数学模型描述:

$$\begin{aligned} \max \quad & z = 7x_1 + 12x_2 \\ \text{s. t.} \quad & \begin{cases} 9x_1 + 4x_2 \leq 360 \\ 4x_1 + 5x_2 \leq 200 \\ 3x_1 + 10x_2 \leq 300 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned} \quad (2.2)$$

式(2.2)表示在所有满足约束条件式(2.1)的可行方案 $[x_1 \ x_2]^T$ 中, 去寻找收入 z 最大的方案 $[x_1^* \ x_2^*]^T$ 。其中的 s. t. 是约束条件的符号, 也可以省略。式(2.2)称为线性规划模型, 它由三个基本要素组成: ①决策变量 $X = [x_1 \ x_2]^T$, 表示一个方案; ②约束条件式(2.1), 表示决策变量必须满足的条件; ③目标函数 $z = 7x_1 + 12x_2$, 表示决策变量 $[x_1 \ x_2]^T$ 优劣的数量指标。

由于式(2.2)中目标函数与约束条件均是线性的, 所以称此模型为线性规划模型。

例 2.1 某化学试剂公司生产两种显影液, 这两种产品每生产 1 L 的成本都为 3 元。根据订货合同, 在二周内化学试剂公司至少要生产出 30 L 的 A 产品和 20 L 的 B 产品。生产这两种显影液都需要一种易变质原料。这种原料现在库存量为 80 kg, 若两周内不使用则变质。每生产 1 L 的 A 产品需 1 kg 此原料; 每生产 1 L 的 B 产品需 2 kg 此原料。公司管理部门需要确定一个两周内成本最低的生产计划, 它至少要用尽库存 80 kg 的易变质原料, 必要时还可再订购此原料, 而且至少生产出 30 L 的 A 产品和 20 L 的 B 产品。

解 用线性规划模型求解。设 x_1 为生产 A 产品的数量; x_2 为生产 B 产品的数量。线性规划模型为

$$\begin{aligned} \min \quad & z = 3x_1 + 3x_2 \quad (\text{成本}) \\ \begin{cases} x_1 + 2x_2 \geq 80 & (\text{原料}) \\ x_1 \geq 30 & (\text{A 产量}) \\ x_2 \geq 20 & (\text{B 产量}) \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

例 2.2 某大学计划早春时期在校园内草坪上施肥, 草坪需要的氮、磷、钾的最低数量以

及市场上销售的三种肥料的成份和价格如表 2.2 所示。

表 2.2

(每 1 000 kg)

元素	最低数量(kg)	肥料	氮含量(kg)	磷含量(kg)	钾含量(kg)	价格(元)
氮	10	I	25	10	5	10
磷	7	II	10	5	10	8
钾	5	III	5	10	5	7

这所大学可以根据需要不受限制地购买到各种肥料,混合后施放到草坪上,列出一个线性规划模型确定购买各种肥料的数量,使之总成本最低。

解 设 x_1 为购买肥料 I 的数量, x_2 为购买肥料 II 的数量, x_3 为购买肥料 III 的数量,则线性规划模型为

$$\begin{aligned} \min z &= \frac{10}{1\,000}x_1 + \frac{8}{1\,000}x_2 + \frac{7}{1\,000}x_3 \\ &\begin{cases} \frac{25}{1\,000}x_1 + \frac{10}{1\,000}x_2 + \frac{5}{1\,000}x_3 \geq 10 \\ \frac{10}{1\,000}x_1 + \frac{5}{1\,000}x_2 + \frac{10}{1\,000}x_3 \geq 7 \\ \frac{5}{1\,000}x_1 + \frac{10}{1\,000}x_2 + \frac{5}{1\,000}x_3 \geq 5 \\ x_1, x_2, x_3 \geq 0 \end{cases} \end{aligned}$$

2.1.2 线性规划的一般模型与标准型

对于一般线性规划模型,目标函数可以求最大(如求利润最大),也可以求最小(如求成本最小),约束条件可以是“ \leq ”,也可以是“ \geq ”或“ $=$ ”型的。因此,一般线性规划模型可表示为

$$\begin{aligned} \max(\min) z &= c_1x_1 + c_2x_2 + \cdots + c_nx_n \\ &\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq (\geq, =) b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq (\geq, =) b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq (\geq, =) b_m \\ x_j \geq 0, j=1, 2, \cdots, n \end{cases} \end{aligned}$$

式中, $X = [x_1 \ x_2 \ \cdots \ x_n]^T$ 为决策变量; $z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$ 为目标函数; $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq (\geq, =) b_i$ 为资源约束条件($i=1, 2, \cdots, m$); $x_j \geq 0 (j=1, 2, \cdots, n)$ 为非负约束条件,称满足全部约束条件的决策变量为可行解,使目标函数为最大(或最小)的可行解称为最优解。

为了论述方便,把最大化、等式约束型的线性规划称为线性规划的标准型,即

$$\begin{aligned} \max z &= c_1x_1 + c_2x_2 + \cdots + c_nx_n \\ &\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\ x_j \geq 0, j=1, 2, \cdots, n \end{cases} \end{aligned} \quad (2.3)$$

标准型的矩阵形式为

$$\max z = CX$$

$$\begin{cases} AX = b \\ X \geq 0 \end{cases}$$

其中 $C = [c_1 \ c_2 \ \cdots \ c_n]$, $X = [x_1 \ x_2 \ \cdots \ x_n]^T$, $b = [b_1 \ b_2 \ \cdots \ b_m]^T$,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

习惯上称 C 为价格系数向量; b 为资源向量, 标准型中设 $b \geq 0$; A 为技术系数矩阵。

对于各种非标准形式的线性规划都可以通过适当的变换化为等价的标准型问题。具体做法如下:

①原问题目标函数求极小, 可以改变目标函数的符号, 然后求极大。例如, 原问题

$$\min z = 2x_1 + 4x_2$$

可化为

$$\max z' = -2x_1 - 4x_2$$

②对“ \leq ”约束, 可添加松弛变量构成新的等式约束, 例如, 对约束 $9x_1 + 4x_2 \leq 360$, 添加松弛变量 $x_3 \geq 0$, 构成等式约束

$$9x_1 + 4x_2 + x_3 = 360$$

目标函数中松弛变量 x_3 的价格系数 $c_3 = 0$ 。

③对“ \geq ”约束, 可减去一个剩余变量构成等式约束, 例如, 对约束 $5x_1 - 2x_2 \geq 2$, 减去剩余变量 $x_4 \geq 0$, 构成等式约束

$$5x_1 - 2x_2 - x_4 = 2$$

同样, 剩余变量的价格系数 $c_4 = 0$ 。相对于松弛或剩余变量, 原有的变量也可称为结构变量。

④对原问题中所含的 x_j 不限制取非负值时称之为自由变量, 并可进行变换

$$x_j = x'_j - x''_j, x'_j \geq 0, x''_j \geq 0$$

代入原问题的目标函数及约束条件中, 化成不含自由变量的线性规划问题。

例 2.3 对线性规划式(2.2), 添加松弛变量 $x_3 \geq 0, x_4 \geq 0, x_5 \geq 0$, 相应 $c_3 = c_4 = c_5 = 0$, 得标准型

$$\begin{aligned} \max z &= 7x_1 + 12x_2 \\ \begin{cases} 9x_1 + 4x_2 + x_3 & = 360 \\ 4x_1 + 5x_2 + x_4 & = 200 \\ 3x_1 + 10x_2 + x_5 & = 300 \\ x_j \geq 0 \quad j=1, 2, \dots, 5 \end{cases} \end{aligned} \quad (2.4)$$

标准型式(2.4)与原问题式(2.2)是等价的。在此例中不难理解, 松弛变量的实际意义是剩余的资源。

例 2.4 将下面非标准线性规划化为等价的标准型线性规划:

$$\min x = -x_1 + 3x_2 - 7x_3$$

$$\begin{cases} x_1 - 2x_2 + 3x_3 \leq 7 \\ 2x_1 - x_2 + x_3 \geq 4 \\ -5x_1 + x_2 + 4x_3 = 6 \\ x_1, x_2 \geq 0, x_3 \text{ 为自由变量} \end{cases}$$

解 按下述步骤进行:

①将目标函数转化为求极大型,即得

$$\max z' = x_1 - 3x_2 + 7x_3$$

②对第一个约束添加松弛变量 $x_4 \geq 0$,得

$$x_1 - 2x_2 + 3x_3 + x_4 = 7$$

③对第二个约束减去剩余变量 $x_5 \geq 0$,得

$$2x_1 - x_2 + x_3 - x_5 = 4$$

④对自由变量 x_3 ,令

$$x_3 = x'_3 - x''_3, x'_3 \geq 0, x''_3 \geq 0$$

原规划化为标准型:

$$\begin{aligned} \max z' &= x_1 - 3x_2 + 7x'_3 - 7x''_3 \\ \begin{cases} x_1 - 2x_2 + 3x'_3 - 3x''_3 + x_4 &= 7 \\ 2x_1 - x_2 + x'_3 - x''_3 - x_5 &= 4 \\ -5x_1 + x_2 + 4x'_3 - 4x''_3 &= 6 \\ x_1, x_2, x'_3, x''_3, x_4, x_5 &\geq 0 \end{cases} \end{aligned}$$

2.1.3 线性规划的图解法

对于只有两个变量的线性规划问题,可以直接用图解法求解。图解法比较直观,而且对一般问题的解决很有启发作用。

图解法可分为两步进行:第一步,根据约束条件画出与约束条件相应方程的直线,由这些直线共同确定的区域即为可行解的区域(满足约束条件的决策变量集合);第二步,画出目标函数的等值线,然后平行移动至与可行区域边界“相切”之点,此点即为最优点,相应坐标 $[x_1, x_2]^T$ 即为最优解。

以问题1为例说明,式(2.2)的约束中每个不等式都表示一个半平面。第一步画出这些半平面的分界直线,即

$$9x_1 + 4x_2 = 360$$

$$4x_1 + 5x_2 = 200$$

$$3x_1 + 10x_2 = 300$$

确定它们与 $x_1 \geq 0, x_2 \geq 0$ 共同围成的部分即为可行域(即画影线部分),如图2.1所示。

第二步作目标函数等值线 $7x_1 + 12x_2 = k$,沿 k 增大方向平行移动,与可行域“相切”于 D 点,故点 D 为最优点。求解两直线之交点,即解方程组

$$\begin{cases} 4x_1 + 5x_2 = 200 \\ 3x_1 + 10x_2 = 300 \end{cases}$$

得 D 的坐标为 $x_1 = 20, x_2 = 24$, 即 $X^* = [20$
 $24]^T$ 为最优生产方案, 最大收入为

$$z^* = (7 \times 20) + (12 \times 24) = 428$$

用图解法求解线性规划时, 可能会出现下面 4 种情况:

①有惟一最优解(图 2.1), 目标函数等值线“相切”于约束集合的一个角点 D , 则 D 点是线性规划问题的惟一最优解;

②有多个最优解(图 2.2), 目标函数等值线平行于一条约束边界, 则此边上所有点都是最优解;

③有可行解, 但没有使目标值为有限的最优解; 无有限最优解, 也称无界解(图 2.3)

④无可行解(图 2.4), 约束集合是空集, 即无可行解。

例 2.5 (与图 2.2 对应)

$$\max z = 2x_1 + 2x_2$$

$$\begin{cases} x_1 + x_2 \leq 3 \\ -x_1 + x_2 \leq 1 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

例 2.6 (与图 2.3 对应)

$$\max z = x_1 + x_2$$

$$\begin{cases} -2x_1 + x_2 \leq 4 \\ x_1 - x_2 \leq 2 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

例 2.7 (与图 2.4 对应)

$$\max z = x_1 + 2x_2$$

$$\begin{cases} x_1 + x_2 \leq 1 \\ x_1 - 2x_2 \geq 2 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

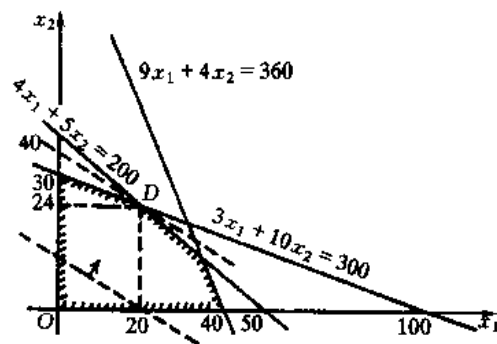


图 2.1

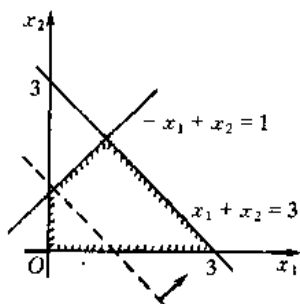


图 2.2

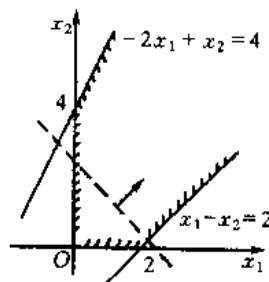


图 2.3

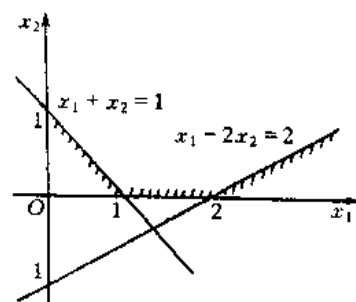


图 2.4

图解法只能用于两个变量的情况, 但从图解法能得到两个重要结论:

①线性规划的约束集合是凸多面体;

②线性规划若有最优解,则最优解一定能在凸多面体的角点(顶点)上达到。

这样,问题就转化为从有限多个角点中寻找最优点,使原来从所有可行解中寻找最优解的工作大大简化。线性规划的单纯形解法的依据就是这两个结论。

2.2 单纯形法

2.2.1 单纯形法的原理与基本概念

单纯形法(simplex method)是由美国数学家 G. B. Dantzig 于 1947 年提出的,到目前为止,它仍是线性规划问题最有效的解法。单纯形法是一种迭代算法。它利用线性规划最优解在可行域角点上得到这一结论,首先确定一个初始角点,用某种法则判别它是否是最优解。若不是最优解,则设法去寻找一个更好的角点。所谓“更好”的角点是指新角点的目标函数值比旧角点的目标函数值大。由于角点个数是有限的,经过有限次迭代(寻找更好角点)后必定达到最优点。这样,单纯形法就需要解决四个问题:①可行域的角点有什么特征? ②如何确定第一个角点? ③如何判定一个角点是否是最优? ④如何从一个角点出发去寻找更好的角点?

首先考虑式(2.3)所定义的线性规划问题。假设 $n \geq m$, 且系数矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

的秩为 m , 用 P_j 表示 A 中第 j 列的列向量, 即

$$P_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

由此, 矩阵 A 可表示为 $A = [P_1 \ P_2 \ \cdots \ P_n]$, 同样有

$$AX = \sum_{j=1}^n x_j P_j$$

为了表达角点的坐标, 引入基、基变量、基本解、基本可行解几个重要概念。

1. 基与基变量

若 $P_{i1}, P_{i2}, \dots, P_{im}$ 是 $A = [P_1 \ P_2 \ \cdots \ P_n]$ 中 m 个线性无关的列, 则称方阵

$$B = [P_{i1} \ P_{i2} \ \cdots \ P_{im}]$$

为一个基, 称其余列构成的矩阵为非基向量矩阵, 记为 N 。

基 B 中各列对应的变量称为基变量, 记为

$$X_B = [x_{i1} \ x_{i2} \ \cdots \ x_{im}]^T$$

其余对应于非基矩阵各列的变量称为非基变量, 记为 X_N 。例如, 在标准型问题式(2.4)中有

$$A = \begin{bmatrix} 9 & 4 & 1 & 0 & 0 \\ 4 & 5 & 0 & 1 & 0 \\ 3 & 10 & 0 & 0 & 1 \end{bmatrix} = [P_1 \ P_2 \ P_3 \ P_4 \ P_5]$$

若取基 $B_1 = [P_3 \ P_4 \ P_5]$, $B_2 = [P_3 \ P_1 \ P_2]$, 则有

$$B_1 = [P_3 \ P_4 \ P_5] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, N_1 = [P_1 \ P_2] = \begin{bmatrix} 9 & 4 \\ 4 & 5 \\ 3 & 10 \end{bmatrix}$$

$$X_{B_1} = [x_3 \ x_4 \ x_5]^T, X_{N_1} = [x_1 \ x_2]^T$$

和

$$B_2 = [P_3 \ P_1 \ P_2] = \begin{bmatrix} 1 & 9 & 4 \\ 0 & 4 & 5 \\ 0 & 3 & 10 \end{bmatrix}, N_2 = [P_4 \ P_5] = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$X_{B_2} = [x_3 \ x_1 \ x_2]^T, X_{N_2} = [x_4 \ x_5]^T$$

2. 基本解与基本可行解

对于确定的一个基 B , 则能由这基 B 去确定约束方程 $AX = b$ 的一个解。为了讲述方便, 假设基 B 是选择由 A 的前 m 列组成(不失一般性), 即有

$$B = [P_1 \ P_2 \ \cdots \ P_m], N = [P_{m+1} \ P_{m+2} \ \cdots \ P_n]$$

约束方程 $AX = b$ 可表示为

$$[B \ N] \begin{bmatrix} X_B \\ X_N \end{bmatrix} = b$$

其中 $A = [B \ N]$, $X = \begin{bmatrix} X_B \\ X_N \end{bmatrix}$, 进一步有

$$BX_B + NX_N = b$$

$$X_B = B^{-1}b - B^{-1}NX_N$$

(2.5)

令非基变量 $X_N = [x_{m+1} \ x_{m+2} \ \cdots \ x_n]^T = 0$, 有

$$X_B = B^{-1}b = \begin{bmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_m \end{bmatrix}$$

这样, 由基 B 确定了方程 $AX = b$ 的一个解

$$X = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} = [b'_1 \ b'_2 \ \cdots \ b'_m \ 0 \ \cdots \ 0]^T$$

称这个解为一个基本解。若它进一步满足非负约束 $x_j \geq 0 (j=1, 2, \cdots, n)$, 即满足

$$B^{-1}b = [b'_1 \ b'_2 \ \cdots \ b'_m]^T \geq 0$$

则称这个解是一个基本可行解, 称基 B 为可行基。

例如, 在标准型式(2.4)中, 取基 $B = (P_3, P_1, P_2)$ 时, 则有相应的基本可行解

$$X_B = [x_3 \quad x_1 \quad x_2]^T = B^{-1}b = \begin{bmatrix} 1 & 9 & 4 \\ 0 & 4 & 5 \\ 0 & 3 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 360 \\ 200 \\ 300 \end{bmatrix} = [84 \quad 20 \quad 24]^T$$

$$X_N = [x_4 \quad x_5]^T = [0 \quad 0]^T$$

即为

$$X = [20 \quad 24 \quad 84 \quad 0 \quad 0]^T$$

可以证明,线性规划可行域角点的坐标恰好是某一个基本可行解,即角点与基本可行解是一一对应的,在角点中寻找最优点即可转化为在所有基本可行解中寻找最优解。因此,只需考虑所有基本可行解就够了。

2.2.2 单纯形法的步骤

单纯形法是一个迭代算法,它由三步组成,见图 2.5。

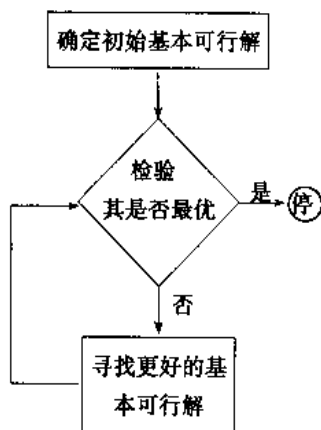


图 2.5

1. 确定一个初始基本可行解

先对原问题添加松弛变量化为标准型,即

$$\max z = CX$$

$$AX = b$$

$$X \geq 0$$

若系数矩阵 A 中含单位矩阵 I ,则取初始基 $B_0 = I$,由此初始基 B_0 可得初始基本可行解

$$X_{B_0} = B_0^{-1}b = b', X_{N_0} = 0$$

若 A 中不含单位矩阵 I ,则可由以后介绍的大 M 法确定一个初始基本可行解。

2. 检验一个基本可行解是否最优

对于给定的可行基 B ,相应基本可行解为 $X_B = B^{-1}b, X_N = 0$ 。而由式(2.5)知,基变量 X_B 可由非基变量 X_N 线性表示为

$$X_B = B^{-1}b - B^{-1}NX_N$$

代入目标函数 $z = CX = C_B X_B + C_N X_N$ 得

$$z = C_B B^{-1}b + (C_N - C_B B^{-1}N)X_N \quad (2.6)$$

其中 C_B 为基变量 $X_B = [x_{B_1} \quad x_{B_2} \quad \cdots \quad x_{B_m}]^T$ 在目标函数中的系数,即

$$C_B = [c_{B_1} \quad c_{B_2} \quad \cdots \quad c_{B_m}]$$

而 C_N 为由非基变量 X_N 在目标函数中系数构成的向量。

从式(2.6)可知,当向量 $C_N - C_B B^{-1}N \leq 0$ 时,对应于这个基本可行解 $X_B = B^{-1}b = [b'_{B_1} \quad b'_{B_2} \quad \cdots \quad b'_{B_m}]^T, X_N = 0$ 的目标函数值 $z = C_B B^{-1}b$ 为最大,即这个基本可行解为最优解。向量 $C_N - C_B B^{-1}N$ 为判别是否最优的检验向量。因此,对给定的一个可行基 B (即给定一个基本可行解 $X_B = B^{-1}b, X_N = 0$),判别它是否最优,只需计算每一非基变量 x_j 的检验数(检验向量 $C_N - C_B B^{-1}N$ 的每一分量)

$$\sigma_j = c_j - C_B B^{-1}P_j$$

式中, c_j 为非基变量 x_j 的价格系数; P_j 为 A 中对应于非基变量 x_j 之列向量。若所有 $\sigma_j \leq 0$ 时, 这个基本可行解为最优, 反之, 若有某一检验数 $\sigma_k > 0$, 则此解一定不是最优。同时可以证明, 若某 $\sigma_k > 0$, 而相应 $B^{-1}P_k \leq 0$ (每个分量 ≤ 0) 时, 原问题没有有限最优解。若讨论的线性规划是 \min 型时, 则当所有 $\sigma_j \geq 0$ 时为最优。

3. 寻找一个更好的基本可行解

对于一个基本可行解 $X_B = B^{-1}b, X_N = 0$, 经第二步检验不是最优解, 则需寻找一个更好的基本可行解, 即去寻找一个新的可行基 \bar{B} 。 \bar{B} 的选择是在基 B 的基础上进行的。假设 $B = [P_1 \cdots P_l \cdots P_m], N = [P_{m+1} \cdots P_n]$, 新可行基 \bar{B} 是由从原非基阵 N 中选择一列 P_k 去代换原基 B 中的某一列 P_l 而得, 即

$$\bar{B} = [P_1 \cdots P_k \cdots P_m]$$

新基 \bar{B} 与旧基 B 仅有一列不同, 问题是如何选择进基列 P_k 和出基列 P_l , 使得构成的新基 \bar{B} 是一个可行基, 而且相应基本可行解的目标函数值上升。

首先决定进基列 P_k , 计算所有非基变量的检验数

$$\max_j \{ \sigma_j > 0 \} = \sigma_k$$

则选择相应的列 P_k 作为进基列。

然后选择出基列 P_l , 计算

$$\theta_l = \min_i \left[\frac{(B^{-1}b)_i}{(B^{-1}P_k)_i} \mid (B^{-1}P_k)_i > 0 \right] = \frac{(B^{-1}b)_l}{(B^{-1}P_k)_l} = \frac{b'_l}{a'_{lk}}$$

$$\text{式中, } B^{-1}b = \begin{bmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_m \end{bmatrix}, B^{-1}P_k = \begin{bmatrix} a'_{1k} \\ a'_{2k} \\ \vdots \\ a'_{mk} \end{bmatrix}。这样选择的出基列为 P_l。$$

可以证明, 依这样的法则得到的新基

$$\bar{B} = [P_1 \cdots P_k \cdots P_m]$$

是一个可行基, 而且相应的目标函数值上升。其新基本可行解 $X_{\bar{B}} = \bar{B}^{-1}b, X_{\bar{N}} = 0$, 转入第二步检验。

2.2.3 单纯形表

上述计算步骤可以通过制作单纯形表在表上计算完成。对每个确定的可行基 B 可作出一个单纯形表, 如表 2.3 所示。

表 2.3

C_B	X_B	$B^{-1}b$	c_1	c_2	\cdots	c_n
			x_1	x_2	\cdots	x_n
c_{B_1}	x_{B_1}	b'_1	a'_{11}	a'_{12}	\cdots	a'_{1n}
c_{B_2}	x_{B_2}	b'_2	a'_{21}	a'_{22}	\cdots	a'_{2n}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
c_{B_m}	x_{B_m}	b'_m	a'_{m1}	a'_{m2}	\cdots	a'_{mn}
检验数 $c_j - C_B B^{-1}P_j$			σ_1	σ_2	\cdots	σ_n

单纯形表中的表头 X_B 下填确定的基变量名, C_B 下填相应的基变量系数, $B^{-1}b$ 下填基本可行解中基变量之解值(非基变量为零)。

$$X_B = B^{-1}b = [b'_1 \quad b'_2 \quad \cdots \quad b'_m]^T$$

x_j 下系数列为

$$P'_j = B^{-1}P_j = B^{-1} \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} = \begin{bmatrix} a'_{1j} \\ a'_{2j} \\ \vdots \\ a'_{mj} \end{bmatrix}$$

最后一行填写检验数

$$\sigma_j = c_j - C_B B^{-1}P_j$$

$$= c_j - [c_{B_1} \quad c_{B_2} \quad \cdots \quad c_{B_m}] \begin{bmatrix} a'_{1j} \\ a'_{2j} \\ \vdots \\ a'_{mj} \end{bmatrix}$$

显然,按上式计算,基变量之检验数必为零。

在每个基的单纯形表上能完成三件事:

- ①确定对应于基的基本可行解,即 $X_B = B^{-1}b$;
- ②计算检验数行值,判断当前解是否最优;
- ③可由表上确定进基变量 x_k 、出基变量 x_l (确定主元素 a'_{lk}),从而确定一个新的更好的基。

由步骤 1 知,用单纯形表计算时,取单位矩阵为初始基 $B_0 = I$,则相应初始单纯形表不必计算,表中各元素就是原始数据,即有

$$B^{-1}b = Ib = b, P'_j = B^{-1}P_j = IP_j = P_j$$

只需计算检验数 $\sigma_j = c_j - C_B B^{-1}P_j$,就建立了初始单纯形表。进一步,再在表上确定相应基本可行解、检验数和新的基 B_1 (确定进基变量 x_k ,出基变量 x_l ,表上元素 a'_{lk} 为主元)。

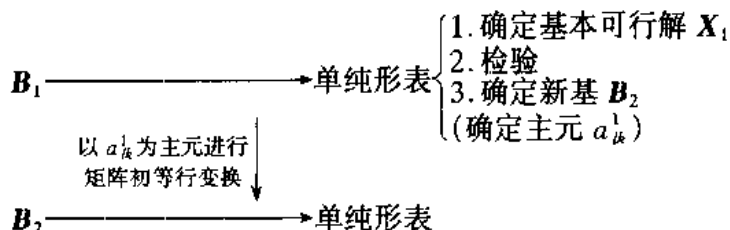
这样确定新基 B_1 后,再去建立 B_1 的单纯形表,在表上重复完成上述三个步骤。在确定新表时,不必先计算逆矩阵 B_1^{-1} ,再分别与 b 及各行 P_j 相乘,而是可把每一单纯形表看成一个矩阵,并可证明, B_1 的单纯形表可由 B_0 的单纯形表进行以 a'_{lk} 为主元的矩阵初等行变换而得,即为

$$\begin{bmatrix} b'_1 & a'_{11} & \cdots & a'_{1k} & \cdots & a'_{1n} \\ \vdots & \vdots & & \vdots & & \vdots \\ b'_l & a'_{l1} & \cdots & a'_{lk} & \cdots & a'_{ln} \\ \vdots & \vdots & & \vdots & & \vdots \\ b'_m & a'_{m1} & \cdots & a'_{mk} & \cdots & a'_{mn} \end{bmatrix} \xrightarrow[\text{矩阵初等行变换}]{\text{以 } a'_{lk} \text{ 为主元}} \begin{bmatrix} b''_1 & a''_{11} & \cdots & 0 & \cdots & a''_{1n} \\ \vdots & \vdots & & \vdots & & \vdots \\ b''_l & a''_{l1} & \cdots & 1 & \cdots & a''_{ln} \\ \vdots & \vdots & & \vdots & & \vdots \\ b''_m & a''_{m1} & \cdots & 0 & \cdots & a''_{mn} \end{bmatrix}$$

整个单纯形表计算过程如下:

$$B_0 = I \xrightarrow{(B^{-1}b=b) \quad (B^{-1}P_j=P_j)} \text{单纯形表} \begin{cases} 1. \text{确定基本可行解 } X_0 \\ 2. \text{检验} \\ 3. \text{确定新基 } B_1 \\ \quad (\text{确定主元 } a'_{lk}) \end{cases}$$

以 a'_{lk} 为主元进行
矩阵初等行变换



以问题 1 标准型式(2.4)为例,用单纯形法求解,式(2.4)中系数矩阵为

$$A = [P_1 \ P_2 \ P_3 \ P_4 \ P_5] = \begin{bmatrix} 9 & 4 & 1 & 0 & 0 \\ 4 & 5 & 0 & 1 & 0 \\ 3 & 10 & 0 & 0 & 1 \end{bmatrix}$$

根据算法,第一步选择单位矩阵为初始基,即有

$$B_0 = [P_3 \ P_4 \ P_5] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

基 B_0 对应的初始单纯形表中元素为

$$P'_j = B_0^{-1} P_j = P_j$$

$$b' = B_0^{-1} b = I^{-1} b = [360 \ 200 \ 300]^T$$

取 $X_{B_0} = [x_3 \ x_4 \ x_5]^T$, $C_{B_0} = [0 \ 0 \ 0]$,即可建立初始单纯形表,见表 2.4。进一步在表上确定:

表 2.4

C_B	X_B	$B^{-1}b$	7 x_1	12 x_2	0 x_3	0 x_4	0 x_5	θ
0	x_3	360	9	4	1	0	0	90
0	x_4	200	4	5	0	1	0	40
0	x_5	300	3	[10]	0	0	1	30
检验数 σ_j			7	12	0	0	0	

①对应于基 B_0 的基本可行解

$$X_{B_0} = [x_3 \ x_4 \ x_5]^T = B^{-1}b = [360 \ 200 \ 300]^T, X_{N_0} = [x_1 \ x_2]^T = [0 \ 0]^T$$

即有基本可行解

$$X_0 = [0 \ 0 \ 360 \ 200 \ 300]^T$$

②从公式 $\sigma_j = c_j - C_B B^{-1} P_j$ 计算各变量之检验数,基变量之检验数 $\sigma_3 = \sigma_4 = \sigma_5 = 0$,不必计算,而非基变量的检验数

$$\sigma_1 = c_1 - C_B B^{-1} P_1 = 7 - [0 \ 0 \ 0] \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix} = 7$$

$$\sigma_2 = c_2 - C_B B^{-1} P_2 = 12 - [0 \ 0 \ 0] \begin{bmatrix} 4 \\ 5 \\ 10 \end{bmatrix} = 12$$

填入表最后一行相应位置,由于检验数不全部小于等于零,基本可行解 X_0 不是最优解;

③计算 $\max\{\sigma_1, \sigma_2\} = \sigma_2 = 12$, 确定 x_2 为进基变量。

计算 $\theta = \min\{b'_i/a'_{i2} | a'_{i2} > 0\} = \min\{360/4, 200/5, 300/10\} = 300/10$, 得 $l = 3$ 。确定第 2 列与第 3 行交叉元素 a'_{32} 为主元, 表示确定 x_2 为进基变量, 基 $B_0 = [P_3 \ P_4 \ P_5]$ 中第 3 列出基, 即 x_5 为出基变量, 从而确定一新基

$$B_1 = [P_3 \ P_4 \ P_2]$$

建立对应于 B_1 的单纯形表, 取 $X_{B_1} = [x_3 \ x_4 \ x_2]^T$, $C_{B_1} = [0 \ 0 \ 12]$ 。表中其他元素可由表 2.2 所构成的矩阵进行以 a'_{32} 为主元的初等行变换而得, 即将主元化为 1, 含主元列的其余元素化为零。即有

$$\begin{bmatrix} 360 & 9 & 4 & 1 & 0 & 0 \\ 200 & 4 & 5 & 0 & 1 & 0 \\ 300 & 3 & 10 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 360 & 9 & 4 & 1 & 0 & 0 \\ 200 & 4 & 5 & 0 & 1 & 0 \\ 30 & 0.3 & 1 & 0 & 0 & 0.1 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 360 & 9 & 4 & 1 & 0 & 0 \\ 50 & 2.5 & 0 & 0 & 1 & -0.5 \\ 30 & 0.3 & 1 & 0 & 0 & 0.1 \end{bmatrix} \rightarrow \begin{bmatrix} 240 & 7.8 & 0 & 1 & 0 & -0.4 \\ 50 & 2.5 & 0 & 0 & 1 & -0.5 \\ 30 & 0.3 & 1 & 0 & 0 & 0.1 \end{bmatrix}$$

这样就建立了对应于基 B_1 的单纯形表, 见表 2.5。

表 2.5

C_B	X_B	$B^{-1}b$	7 x_1	12 x_2	0 x_3	0 x_4	0 x_5	θ
0	x_3	240	7.8	0	1	0	-0.4	30.8
0	x_4	50	[2.5]	0	0	1	-0.5	20
12	x_2	30	0.3	1	0	0	0.1	100
检验数 σ_j			3.4	0	0	0	-1.2	

在单纯形表上确定:

①对应于基 $B_1 = [P_3 \ P_4 \ P_2]$ 的基本可行解为

$$X_1 = [0 \ 30 \ 240 \ 50 \ 0]^T$$

②计算检验数, 其中基变量下检验数均为零, 即 $\sigma_3 = \sigma_4 = \sigma_2 = 0$, 非基变量下检验数为

$$\sigma_1 = c_1 - C_{B_1} B_1^{-1} P_1 = 7 - [0 \ 0 \ 12] \begin{bmatrix} 7.8 \\ 2.5 \\ 0.3 \end{bmatrix} = 3.4$$

$$\sigma_5 = c_5 - C_{B_1} B_1^{-1} P_5 = 0 - [0 \ 0 \ 12] \begin{bmatrix} -0.4 \\ -0.5 \\ 0.1 \end{bmatrix} = -1.2$$

填在表中最后一行相应位置, 由于 $\sigma_1 > 0$, 所以当前基本可行解 X_1 不是最优;

③计算 $\max\{\sigma_j > 0\} = \sigma_1 = 3.4$, 选择 x_1 为进基变量(或说 P_1 列为进基列), 计算 $\theta = \min\{240/7.8, 50/2.5, 30/0.3\} = b'_i/a'_{i1} = 50/2.5$, 得 $l = 2$;

④确定主元为第 1 列与第 2 行交叉元素 a'_{21} , 即由 P_1 列进基, 基 $B_1 = [P_3 \ P_4 \ P_2]$ 中第 2 列 P_4 出基(x_4 为出基变量), 即由 P_1 代替 P_4 而得新基

$$B_2 = [P_3 \ P_1 \ P_2]$$

同样建立基 B_2 的单纯形表, 取 $X_{B_2} = [x_3 \ x_1 \ x_2]^T$, $C_{B_2} = [0 \ 7 \ 12]$ 。单纯形表中其他元素可由表 2.3 构成的矩阵进行以 a'_{21} 为主元的初等行变换得到, 即有

$$\begin{bmatrix} 240 & 7.8 & 0 & 1 & 0 & -0.4 \\ 50 & (2.5) & 0 & 0 & 1 & -0.5 \\ 30 & 0.3 & 1 & 0 & 0 & 0.1 \end{bmatrix} \rightarrow \begin{bmatrix} 84 & 0 & 0 & 1 & -3.12 & 1.16 \\ 20 & 1 & 0 & 0 & 0.4 & -0.2 \\ 24 & 0 & 1 & 0 & -0.12 & 0.16 \end{bmatrix}$$

这样就能建立对应于基 $B_2 = [P_3 \ P_1 \ P_2]$ 的单纯形表, 见表 2.6。

表 2.6

C_B	X_B	$B^{-1}b$	7 x_1	12 x_2	0 x_3	0 x_4	0 x_5	θ
0	x_3	84	0	0	1	-3.12	1.16	
7	x_1	20	1	0	0	0.4	-0.2	
12	x_2	24	0	1	0	-0.12	0.16	
检验数 σ_j			0	0	0	-1.36	-0.52	

同样在表上确定:

①对应于 B_1 之基本可行解为

$$X_2 = [20 \ 24 \ 84 \ 0 \ 0]^T$$

②计算检验数。基变量检验数 $\sigma_1 = \sigma_2 = \sigma_3 = 0$, 非基变量检验数

$$\sigma_4 = c_4 - C_{B_2} B_2^{-1} P_4 = 0 - [0 \ 7 \ 12] \begin{bmatrix} -3.12 \\ 0.4 \\ -0.12 \end{bmatrix} = -1.36$$

$$\sigma_5 = c_5 - C_{B_2} B_2^{-1} P_5 = 0 - [0 \ 7 \ 12] \begin{bmatrix} 1.16 \\ -0.2 \\ 0.16 \end{bmatrix} = -0.52$$

由于所有检验数 $\sigma_j \leq 0$, 则当前基本可行解

$$X_2 = [20 \ 24 \ 84 \ 0 \ 0]^T$$

为最优解, 最优值 $z^* = (7 \times 20) + (12 \times 24) = 428$

使用单纯形表求解 min 型问题, 在进行检验时和 max 型相反。当所有检验数 $\sigma_j \geq 0$ 时为最优; 在选择进基变量时, 由

$$\sigma_k = \min\{\sigma_j < 0\}$$

确定 x_k 进基, 其他各步骤与 max 型完全一样, 可用表 2.7 说明。

表 2.7

类型	检验数	判别	确定进基变量 x_k
max	$\sigma_j = c_j - C_B B^{-1} P_j$	所有 $\sigma_j \leq 0$ 最优	$\sigma_k = \max\{\sigma_j > 0\}$
min	$\sigma_j = c_j - C_B B^{-1} P_j$	所有 $\sigma_j \geq 0$ 最优	$\sigma_k = \min\{\sigma_j < 0\}$

例 2.8 用单纯形法求解线性规划问题

$$\max z = 10x_1 + 9x_2 + 4x_3 + 6x_4$$

$$\begin{cases} 3x_1 + 2x_2 + 4x_3 + 2x_4 \leq 70 \\ 5x_1 + 5x_2 + x_3 + 3x_4 \leq 60 \\ 5x_1 + 6x_2 + 3x_3 + x_4 \leq 25 \\ x_i \geq 0, i=1,2,3,4 \end{cases}$$

解 增加松弛变量 x_5, x_6, x_7 , 将上述问题化为标准型:

$$\begin{aligned} \max z &= 10x_1 + 9x_2 + 4x_3 + 6x_4 \\ \begin{cases} 3x_1 + 2x_2 + 4x_3 + 2x_4 + x_5 &= 70 \\ 5x_1 + 5x_2 + x_3 + 3x_4 + x_6 &= 60 \\ 5x_1 + 6x_2 + 3x_3 + x_4 + x_7 &= 25 \\ x_i \geq 0, i=1, \dots, 7 \end{cases} \end{aligned}$$

初始基选 $B_0 = [P_5 \ P_6 \ P_7]$, 解题过程由表 2.8 给出。

表 2.8

C_B	X_B	$B^{-1}b$	10 x_1	9 x_2	4 x_3	6 x_4	0 x_5	0 x_6	0 x_7	θ
0	x_5	70	3	2	4	2	1	0	0	70/3
0	x_6	60	5	5	1	3	0	1	0	60/5
0	x_7	25	[5]	6	3	1	0	0	1	25/5
检验数 σ_j			10↑	9	4	6	0	0	0	
0	x_5	55	0	-8/5	11/5	7/5	1	0	-3/5	275/7
0	x_6	35	0	-1	-2	[2]	0	1	-1	35/2
10	x_1	5	1	6/5	3/5	1/5	0	0	1/5	25
检验数 σ_j			0	-3	-2	4↑	0	0	2	
0	x_5	61/2	0	-9/10	18/5	0	1	-7/10	1/10	305/36
6	x_4	35/2	0	-1/2	-1	1	0	1/2	-1/2	
10	x_1	3/2	1	13/10	[4/5]	0	0	-1/10	3/10	15/8
检验数 σ_j			0	-1	2↑	0	0	-2	0	
0	x_5	95/4	-9/2	-27/4	0	0	1	-1/4	-5/4	
6	x_4	155/8	5/4	9/8	0	1	0	3/8	-1/8	
4	x_3	15/8	5/4	13/8	1	0	0	-1/8	3/8	
检验数 σ_j			-5/2	-17/4	0	0	0	-7/4	-3/4	

最优解 $X^* = [0 \ 0 \ 15/8 \ 155/8 \ 95/4 \ 0 \ 0]^T, z^* = 123 \frac{3}{4}$

2.2.4 大 M 法

上述单纯形法第一步中,若系数矩阵中含有单位矩阵,则取初始基 $B_0 = I$ 。一般若系数矩阵中不含单位矩阵,则可采用大 M 法,通过增加人工变量,构成一个系数矩阵中含有单位矩阵的新的线性规划,然后使用单纯形法求出最优解。这个新的线性规划问题和原问题在一定条件下是等价的。通过下面例子具体说明如何使用大 M 法。

例 2.9 求解线性规划

$$\min z = -4x_1 + 3x_2 + 2x_3$$

$$\begin{cases} x_1 - 2x_2 + 2x_3 \leq 8 \\ -2x_1 + x_2 + x_3 \geq 4 \\ -x_1 + x_3 = 2 \\ x_i \geq 0 \quad (i=1,2,3) \end{cases}$$

解 首先增加松弛变量与剩余变量 x_4, x_5 , 将模型约束化为标准型

$$\begin{cases} x_1 - 2x_2 + 2x_3 + x_4 = 8 \\ -2x_1 + x_2 + x_3 - x_5 = 4 \\ -x_1 + x_3 = 2 \\ x_i \geq 0 \quad (i=1,2,3,4,5) \end{cases}$$

它的系数矩阵为

$$\begin{bmatrix} 1 & -2 & 2 & 1 & 0 \\ -2 & 1 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

此系数矩阵中不含单位矩阵, 在第二、第三个约束方程左边分别添加人工变量 $x_6 \geq 0, x_7 \geq 0$, 即构成一个系数矩阵中含有单位矩阵的新的线性规划问题。在求极大型 \max 问题中, 人工变量在目标函数中系数均为 $-M$; 在求极小型 \min 问题中, 人工变量在目标函数中系数均为 M , 其中 M 是很大很大的正数。上述例题添加人工变量后构成新的线性规划为

$$\begin{cases} \min z = -4x_1 + 3x_2 + 2x_3 + Mx_6 + Mx_7 \\ \begin{cases} x_1 - 2x_2 + 2x_3 + x_4 = 8 \\ -2x_1 + x_2 + x_3 - x_5 + x_6 = 4 \\ -x_1 + x_3 + x_7 = 2 \end{cases} \\ x_i \geq 0 \quad (i=1,2,\dots,7) \end{cases}$$

其中 x_4 是松弛变量, x_5 是剩余变量, x_6, x_7 是人工变量。它的系数矩阵为

$$\bar{A} = \begin{bmatrix} 1 & -2 & 2 & 1 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

由于上述线性规划系数矩阵中含有单位矩阵, 可取 $B_0 = [P_4 \ P_6 \ P_7] = I$, 按上节所述建立单纯形表 2.9。

表 2.9

C_B	X_B	$B^{-1}b$	-4 x_1	3 x_2	2 x_3	0 x_4	0 x_5	M x_6	M x_7	θ
0	x_4	8	1	-2	2	1	0	0	0	4
(a) M	x_6	4	-2	1	1	0	-1	1	0	4
M	x_7	2	-1	0	[1]	0	0	0	1	2
检验数			$-4+3M$	$3-M$	$2-2M$	0	M	0	0	

续表

C_B	X_B	$B^{-1}b$	-4 x_1	3 x_2	2 x_3	0 x_4	0 x_5	M x_6	M x_7	θ
0	x_4	4	3	-2	0	1	0	0	-2	2
(b) M	x_6	2	-1	[1]	0	0	-1	1	-1	
2	x_3	2	-1	0	1	0	0	0	1	
检验数			$-2+M$	$3-M$	0	0	M	0	$2M-2$	
0	x_4	8	1	0	0	1	-2	2	-4	
(c)3	x_2	2	-1	1	0	0	-1	1	-1	
2	x_3	2	-1	0	1	0	0	0	1	
检验数			1	0	0	0	3	$M-3$	$M+1$	

在初始单纯形表 2.9(a) 上可确定:

①基本可行解 $X_0 = [0 \ 0 \ 0 \ 8 \ 0 \ 4 \ 2]^T$;

②非基变量检验数 $\sigma_1 = -4 + 3M, \sigma_2 = 3 - M, \sigma_3 = 2 - 2M, \sigma_5 = M$, 其中 $\sigma_2 < 0, \sigma_3 < 0$, 则 X_0 不是最优解;

③ $\min\{\sigma_2 < 0, \sigma_3 < 0\} = \sigma_3$, 确定 x_3 为进基变量;

④ $\min\{8/2, 4/1, 2/1\} = 2/1$, 得 $l = 3$, 确定基 $B_0 = [P_4 \ P_6 \ P_7]$ 中第三列 P_7 出基(即 x_7 为出基变量, a'_{33} 为主元), 得新基 $B_1 = [P_4 \ P_6 \ P_3]$ 。

建立基 B_1 的单纯形表。 $X_{B_1} = [x_4 \ x_6 \ x_3]^T, C_{B_1} = [0 \ M \ 2]$, 表 2.9(a) 中以 a'_{33} 为主元的初等行变换, 即可得 B_1 的单纯形表 2.9(b)。同样在表上计算出各检验数, 得知基本可行解 $X_1 = [0 \ 0 \ 2 \ 4 \ 0 \ 2 \ 0]^T$ 仍不是最优, 按法则确定表 2.9(b) 中 a'_{22} 为主元, 即得新基 $B_2 = [P_4 \ P_2 \ P_3]$, 同样进行表 2.9(b) 中以 a'_{22} 为主元的初等行变换, 可得 B_2 的单纯形表 2.9(c)。进一步在此表上确定: 基本可行解 $X_2 = [0 \ 2 \ 2 \ 8 \ 0 \ 0 \ 0]^T$, 非基变量检验数 $\sigma_1 = 1 > 0, \sigma_5 = 3 > 0, \sigma_6 = M - 3 > 0, \sigma_7 = M + 1 > 0$, 则基本可行解 $X_2 = [0 \ 2 \ 2 \ 8 \ 0 \ 0 \ 0]^T$ 为新问题的最优解。

使用大 M 法会出现下列两种情况: ①最优解的基变量中含有人工变量, 可以证明此情况下, 原问题无可行解; ②最优解的基变量中不含人工变量, 即人工变量均为零, 可以证明在此情况下, 从最优解中去掉人工变量即为原问题的最优解。

上例中 $X_2 = [0 \ 2 \ 2 \ 8 \ 0]^T$ 即为原问题最优解。

例 2.10 用大 M 法求解线性规划

$$\begin{aligned} \max z &= 5x_1 + 3x_2 + 2x_3 + 4x_4 \\ \begin{cases} 5x_1 + x_2 + x_3 + 8x_4 = 10 \\ 2x_1 + 4x_2 + 3x_3 + 2x_4 = 10 \\ x_i \geq 0 \quad (i=1, 2, 3, 4) \end{cases} \end{aligned}$$

解 系数矩阵中不含单位矩阵, 使用大 M 法增加人工变量 x_5, x_6 , 其目标函数中系数为 $-M$, 得

$$\max z = 5x_1 + 3x_2 + 2x_3 + 4x_4 - Mx_5 - Mx_6$$

$$\begin{cases} 5x_1 + x_2 + x_3 + 8x_4 + x_5 = 10 \\ 2x_1 + 4x_2 + 3x_3 + 2x_4 + x_6 = 10 \\ x_i \geq 0 \quad (i=1,2,\dots,6) \end{cases}$$

取 $B_0 = [P_6 \ P_7]$, 建立单纯形表 2.10 进行计算。

表 2.10

C_B	X_B	$B^{-1}b$	5 x_1	3 x_2	2 x_3	4 x_4	-M x_5	-M x_6	θ
-M	x_5	10	5	1	1	[8]	1	0	10/8
-M	x_6	10	2	4	3	2	0	1	5
检验数			5+7M	3+5M	2+4M	4+10M↑	0	0	
4	x_4	5/4	5/8	1/8	1/8	1		0	10
-M	x_6	15/2	3/4	[15/4]	11/4	0		1	2
检验数			$\frac{5}{2} + \frac{3}{4}M$	$\frac{5}{2} + \frac{15}{4}M↑$	$\frac{3}{2} + \frac{11}{4}M$	0		0	
4	x_4	1	[3/5]	0	1/30	1			5/3
3	x_2	2	1/5	1	11/15	0			10
检验数			2↑	0	-1/3	0			
5	x_1	5/3	1	0	1/18	5/3			
3	x_2	5/3	0	1	13/18	-1/3			
检验数			0	0	-4/9	-10/3			

在使用大 M 法时,可以证明,人工变量被换出基后,不会再被换入基中。所以,若某人工变量被换出基,则这一列不必再参加迭代运算。上例终表中,最优解的基变量不含人工变量,则原问题的最优解为

$$X^* = \left[\frac{5}{3} \quad \frac{5}{3} \quad 0 \quad 0 \right]^T, z^* = \frac{40}{3}$$

例 2.11 求解线性规划

$$\max z = 10x_1 + 15x_2 + 12x_3$$

$$\begin{cases} 5x_1 + 3x_2 + x_3 \leq 9 \\ -5x_1 + 6x_2 + 15x_3 \leq 15 \\ 2x_1 + x_2 + x_3 \geq 5 \\ x_i \geq 0 \quad (i=1,2,3) \end{cases}$$

解 增加松弛变量 x_4, x_5 , 剩余变量 x_6 , 人工变量 x_7 , 求解下述问题:

$$\max z = 10x_1 + 15x_2 + 12x_3 - Mx_7$$

$$\begin{cases} 5x_1 + 3x_2 + x_3 + x_4 = 9 \\ -5x_1 + 6x_2 + 15x_3 + x_5 = 15 \\ 2x_1 + x_2 + x_3 - x_6 + x_7 = 5 \\ x_j \geq 0 \quad (j=1,2,\dots,7) \end{cases}$$

建立单纯形表,计算过程由表 2.11 给出。

表 2.11

C_B	X_B	$B^{-1}b$	10 x_1	15 x_2	12 x_3	0 x_4	0 x_5	0 x_6	$-M$ x_7	θ
0	x_4	9	[5]	3	1	1	0	0	0	9/5
0	x_5	15	-5	6	15	0	1	0	0	
$-M$	x_7	5	2	1	1	0	0	-1	1	5/2
检验数			$10+2M \uparrow$	$15+M$	$12+M$	0	0	$-M$	0	
10	x_1	9/5	1	3/5	1/5	1/5	0	0	0	9
0	x_5	24	0	9	[16]	1	1	0	0	3/2
$-M$	x_7	7/5	0	-1/5	3/5	-2/5	0	-1	1	7/3
检验数			0	$9-\frac{1}{5}M$	$10+\frac{3}{5}M \uparrow$	$-2/5M$	0	$-M$	0	
10	x_1	3/2	1	39/80	0	3/16	-1/80	0	0	
12	x_3	3/2	0	9/16	1	1/16	1/16	0	0	
$-M$	x_7	1/2	0	-43/80	0	-7/16	-3/80	-1	1	
检验数			0	$\frac{27}{8}-\frac{43}{80}M$	0	$-\frac{21}{8}-\frac{7}{16}M$	$-\frac{5}{8}-\frac{3}{80}M$	$-M$	0	

终表最优解基变量中含有人工变量 x_7 , 所以原问题无可行解。

2.2.5 解的几种情况

从单纯形表的终表中能判别线性规划解的几种情况(以 max 型为例)。

1. 惟一最优解

在单纯形终表中, 所有非基变量的检验数 $\sigma_j < 0$ (min 型中, 所有非基变量的检验数 $\sigma_j > 0$), 则此问题只有惟一最优解。如表 2.4 中, 非基变量的检验数 $\sigma_4 = -1.36$ 、 $\sigma_5 = -0.52$ 均严格小于零, 则此问题只有惟一最优解。

2. 多个最优解

在单纯形终表中, 所有检验数 $\sigma_j \leq 0$ (min 型所有 $\sigma_j \geq 0$), 而某一个非基变量 x_k 之检验数 $\sigma_k = 0$ 时, 则原问题有多个最优解。

例 2.12 求解下面的线性规划问题

$$\max z = 2.5x_1 + x_2$$

$$\begin{cases} 3x_1 + 5x_2 \leq 15 \\ 5x_1 + 2x_2 \leq 10 \\ x_1, x_2 \geq 0 \end{cases}$$

解 增加松弛变量 x_3, x_4 , 化为标准型

$$\max z = 2.5x_1 + x_2$$

$$\begin{cases} 3x_1 + 5x_2 + x_3 = 15 \\ 5x_1 + 2x_2 + x_4 = 10 \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

解题过程由表 2.12 给出。

表 2.12

C_B	X_B	$B^{-1}b$	2.5 x_1	1 x_2	0 x_3	0 x_4	θ
0	x_3	15	3	5	1	0	5
0	x_4	10	[5]	2	0	1	2
	σ_j		2.5	1	0	0	
0	x_3	9	0	$\frac{19}{5}$	1	$-\frac{3}{5}$	
2.5	x_1	2	1	$\frac{2}{5}$	0	$\frac{1}{5}$	
	σ_j		0	0	0	$-\frac{1}{2}$	

因非基变量 x_2, x_4 的检验数都小于等于零, 最优解为

$$x_1=2, x_2=0, x_3=9, x_4=0, z^*=5$$

但由于非基变量 x_2 的检验数为 0, 故本题有无穷多个最优解。

3. 无有限最优解

在某单纯形表中, 某一变量的检验数 $\sigma_k > 0$ (若极小问题时为 $\sigma_k < 0$), 而表上相应列元素全小于或等于零, 即

$$P'_k = B^{-1}P_k = \begin{bmatrix} a'_{1k} \\ a'_{2k} \\ \vdots \\ a'_{mk} \end{bmatrix} \leq 0$$

则原问题无有限最优解。

例 2.13 用单纯形法求解下面线性规划问题

$$\max z = 2x_1 + 2x_2$$

$$\begin{cases} -x_1 + x_2 \leq 1 \\ x_1 - 2x_2 \geq -4 \\ x_1, x_2 \geq 0 \end{cases}$$

解 增加松弛变量 x_3, x_4 , 化为标准型

$$\max z = 2x_1 + 2x_2$$

$$\begin{cases} -x_1 + x_2 + x_3 = 1 \\ -x_1 + 2x_2 + x_4 = 4 \\ x_1, x_2, x_3, x_4 \geq 0 \end{cases}$$

解题过程由表 2.13 给出。

表 2.13

C_B	X_B	$B^{-1}b$	2 x_1	2 x_2	0 x_3	0 x_4	θ
0	x_3	1	-1	1	1	0	5
0	x_4	4	-1	2	0	1	2
	σ_j	2	2	0	0	0	

由于当前单纯形表中非基变量 x_1 的检验数 $\sigma_1 = 2 > 0$, 但相应列上的系数均小于零, 则此线性规划问题无有限最优解。

4. 退化解

在单纯形终表中, 最优解为

$$X_B = B^{-1}b = [b'_1 \quad b'_2 \quad \cdots \quad b'_m]^T$$

若其中某分量 $b'_k = 0$, 则称此解为一个退化解。

2.3 对偶问题与灵敏度分析

2.3.1 线性规划的对偶问题

1. 对偶问题的定义

问题 2 对于问题 1, 假设某一公司欲购买这个工厂的全部资源(煤、电、油), 公司提出的价格为 y_1 (煤)、 y_2 (电)、 y_3 (油)。该公司希望确定一组价格, 一方面使工厂能接受这价格而同意出售资源, 另一方面又使公司购买这些资源的总成本最低。这样一个寻求最优价格的问题可由另一个线性规划确定, 即

$$\begin{aligned} \min w &= 360y_1 + 200y_2 + 300y_3 \quad (\text{总成本}) \\ &\begin{cases} 9y_1 + 4y_2 + 3y_3 \geq 7 \\ 4y_1 + 5y_2 + 10y_3 \geq 12 \\ y_1, y_2, y_3 \geq 0 \end{cases} \end{aligned} \quad (2.7)$$

(价格 y_1, y_2, y_3 为厂方能接受的水平)

在这个线性规划中, 决策变量 $[y_1 \quad y_2 \quad y_3]^T$, 分别表示购买煤、电、油之单价, 目标函数为购买工厂中煤、电、油资源的总成本, 而约束条件表示厂方能接受的价格所必须满足的条件, 即厂方按这个价格出售生产单位产品所需资源的收入, 必须不低于工厂用这些资源自己生产产品所得的收入, 只有这样的价格工厂才会接受, 同意出售自己的资源。

例如, 生产 1 单位甲产品耗煤 9 t、电 4 kW·h、油 3 t, 出售这些资源的收入为

$$9y_1 + 4y_2 + 3y_3$$

而用这些资源自己生产, 生产 1 个单位甲产品的收入为 7 万元。所以, 公司提出的价格只有在使得 $9y_1 + 4y_2 + 3y_3 \geq 7$ 时, 工厂才会同意出售。

问题 1 和问题 2 无论从经济意义上或者是在数学形式上都是紧密相连的。从经济上看, 问题 1 是寻找最优生产方案, 以获得最大收入; 问题 2 是寻找最优价格使总成本最低, 从数学模型的形式上看, 它们也是关联的, 比较两模型如下:

问题 1 式(2.2)	问题 2 式(2.7)
价格系数	资源向量
资源向量	价格系数
最大化	最小化
变量个数	约束个数
约束个数	变量个数

约束系数行 \longleftrightarrow 约束系数列

把线性规划(2.2)称为原问题,而把线性规划(2.7)称为其对偶问题。其一般式为

$$\begin{array}{ll} \max & z = CX \\ \text{原问题(P)} & \left\{ \begin{array}{l} AX \leq b \\ X \geq 0 \end{array} \right. \end{array} \quad \begin{array}{ll} \min & w = Yb \\ \text{对偶问题(D)} & \left\{ \begin{array}{l} YA \geq C \\ Y \geq 0 \end{array} \right. \end{array}$$

更一般地,线性规划总可表示为

$$\begin{aligned} \max \quad & z = \sum_{j=1}^n c_j x_j \\ \left\{ \begin{array}{l} \sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i \in I_1) \\ \sum_{j=1}^n a_{ij} x_j = b_i \quad (i \in I_2) \\ x_j \geq 0 \quad (j \in J_1) \end{array} \right. \end{aligned}$$

在这样一个线性规划中, I_1 是不等式约束标号集合, $I_1 \cup I_2 = I = \{1, 2, \dots, m\}$; I_2 是等式约束标号集合; $J_1 \subset J = \{1, 2, \dots, n\}$ 是非负变量标号集合。注意, 上述线性规划中 b_i 是可以取负值的。

它的矩阵形式为

$$\begin{aligned} \max \quad & z = CX \\ \left\{ \begin{array}{l} AX \leq b_{I_1}^1 \\ x_j \geq 0 \quad (j \in J_1) \end{array} \right. \end{aligned}$$

把上述线性规划称为原规划, 其中, $b_{I_1}^1$ 为由所有 $b_i (i \in I_1 \cup I_2)$ 组成的向量, 它的对偶问题定义为下述线性规划:

$$\begin{aligned} \min \quad & w = \sum_{i=1}^m b_i y_i \\ \left\{ \begin{array}{l} \sum_{i=1}^m a_{ij} y_i \geq c_j \quad (j \in J_1) \\ \sum_{i=1}^m a_{ij} y_i = c_j \quad (j \in (J - J_1)) \\ y_i \geq 0 \quad (i \in I_1) \end{array} \right. \end{aligned}$$

式中, J_1 为原规划中非负变量标号集合; $(J - J_1)$ 为自由变量标号集合; I_1 为原规划中“小于等于”约束的标号集合。显然, 对偶问题的矩阵形式可表示为

$$\begin{aligned} \min \quad & w = b^1 Y \\ \left\{ \begin{array}{l} A^T Y \geq C_{J-J_1}^{T_1} \\ y_i \geq 0 \quad (i \in I_1) \end{array} \right. \end{aligned}$$

其中 $C_{J-J_1}^{T_1}$ 表示由所有 $C_j (j \in J_1 \cup (J - J_1))$ 组成的向量。从上述定义看出, 原问题必须表述为约束条件只有“ \leq ”和“ $=$ ”两种形式; 而对偶问题中只有“ \geq ”和“ $=$ ”两种形式; 对偶问题中第 j 行约束究竟是大于还是等于形式, 取决于原问题中第 j 个变量 x_j 是非负还是自由变

量。 x_j 是自由变量则对偶问题中第 j 行是等式约束, x_j 是非负变量则对偶问题中第 j 行为大于等于约束。同样, 对偶变量中变量 y_j 是否是自由变量取决于原问题第 j 行约束是否是等式形式。若原问题第 j 行约束是等式约束, 则 y_j 为自由变量。还可以看出, 原问题中目标函数的系数向量(价格向量)是对偶问题中的约束右边的常数向量(资源向量), 原问题约束中右边常数向量是对偶问题中目标函数的系数向量。

例 2.14 写出下面线性规划的对偶规划:

$$\begin{aligned} \max z &= 5x_1 + x_2 - 3x_3 \\ \begin{cases} 2x_1 & + 2x_2 - x_3 \geq 1 \\ x_1 & - 3x_2 + 4x_3 \leq 10 \\ 2x_1 & + 2x_2 + x_3 = 5 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

解 首先将原问题改为规定形式, 即为

$$\begin{aligned} \max z &= 5x_1 + x_2 - 3x_3 \\ \begin{cases} -2x_1 & - 2x_2 + x_3 \leq -1 \\ x_1 & - 3x_2 + 4x_3 \leq 10 \\ 2x_1 & + 2x_2 + x_3 = 5 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

显然其中 $I_1 = \{1, 2\} \subset I = \{1, 2, 3\}$, $J_1 = \{1, 2\} \subset J = \{1, 2, 3\}$, $(J - J_1) = \{3\}$ 。它的对偶问题为

$$\begin{aligned} \min w &= -y_1 + 10y_2 + 5y_3 \\ \begin{cases} -2y_1 & + y_2 + 2y_3 \geq 5 \\ -2y_1 & - 3y_2 + 2y_3 \geq 1 \\ y_1 & + 4y_2 + y_3 = -3 \\ y_1 \geq 0, y_2 \geq 0 \end{cases} \end{aligned}$$

根据上述定义可归纳为表 2.14 所示关系。

表 2.14

原问题	对偶问题
约束个数 m	变量个数 m
变量个数 n	约束个数 n
目标函数系数向量	约束右边向量
第 j 行约束为等式	第 j 个变量为自由变量
第 j 个变量为自由变量	第 j 行约束为等式
第 j 个变量为非负变量	第 j 行约束为“ \geq ”约束

2. 线性规划对偶问题的基本性质

以下不加证明地给出线性规划对偶问题的几条基本性质。

1) 对称性 一个线性规划的对偶问题的对偶问题恰是原问题。

2) 弱对偶性 假定 X 是原规划(P)的任一可行解, Y 是对偶规划(D)的任一可行解, 则有

$$CX \leq b^T Y$$

3) 无界性 若原问题(对偶问题)为无界解,则其对偶问题(原问题)无可行解(逆命题不成立)。

4) 设 \bar{X} 是原问题的可行解, \bar{Y} 是对偶问题的可行解。当 $C\bar{X} = b^T \bar{Y}$ 时, \bar{X} 、 \bar{Y} 皆为最优解。

5) 强对偶性 原规划有最优解,则对偶规划也有最优解,且最优值相同。

可以证明,(P)和(D)的解一般说来共有下述三种情况:①两个问题都有有限最优解;②两个问题都无可行解;③一个问题有无界解,另一问题无可行解。

6) 互补松弛性 在线性规划问题的最优解中,若对应某一约束条件的对偶变量值为非零,则该约束条件取严格等式;另一方面,如果约束条件取严格不等式,则其对应的对偶变量一定为零。也即若 $\bar{y}_i > 0$, 则 $\sum_{j=1}^n a_{ij} \bar{x}_j = b_i$; 若 $\sum_{j=1}^n a_{ij} \bar{x}_j < b_i$, 则 $\bar{y}_i = 0$ 。

7) 检验数 原问题单纯形表中的检验数对应于对偶问题的一个基本解(由基产生的解,但不满足非负约束)。最终表下的检验数对应于对偶问题的最优解,如表 2.15 所示。

表 2.15

	C	0
$C_B \quad X_B \quad B^{-1}b$	$B^{-1}A$	B^{-1}
σ_j	$C - C_B B^{-1}A$	$-C_B B^{-1}$

其中 $C - C_B B^{-1}A$ 是原问题结构变量下的检验数; $-C_B B^{-1}$ 是松弛变量下的检验数。 $Y = C_B B^{-1}$, $Y_n = C_B B^{-1}A - C$ 是对偶问题最优解的结构变量和松弛变量之解值。

例 2.15 已知线性规划问题

$$\begin{aligned} \min w &= 2x_1 + 3x_2 + 5x_3 + 2x_4 + 3x_5 \\ \begin{cases} x_1 + x_2 + 2x_3 + x_4 + 3x_5 \geq 4 \\ 2x_1 - x_2 + 3x_3 + x_4 + x_5 \geq 3 \\ x_j \geq 0, j=1, 2, \dots, 5 \end{cases} \end{aligned}$$

已知其对偶问题的最优解为 $y_1^* = \frac{4}{5}$, $y_2^* = \frac{3}{5}$, $z = 5$, 试用对偶理论找出原问题的最优解。

解 对偶问题为

$$\begin{aligned} \max z &= 4y_1 + 3y_2 \\ \text{s. t. } \begin{cases} y_1 + 2y_2 \leq 2 & (1) \\ y_1 - y_2 \leq 3 & (2) \\ 2y_1 + 3y_2 \leq 5 & (3) \\ y_1 + y_2 \leq 2 & (4) \\ 3y_1 + y_2 \leq 3 & (5) \\ y_1, y_2 \geq 0 \end{cases} \end{aligned}$$

将 y_1^* 、 y_2^* 的值代入约束条件,得(2)、(3)、(4)为严格不等式;由互补松弛性得 $x_2^* = x_3^* = x_4^* = 0$, 因 $y_1, y_2 > 0$, 则原问题的两个约束条件应取等式,故有

$$x_1^* + 3x_5^* = 4$$

$$2x_1^* + x_5^* = 3$$

求解后得到 $x_1^* = 1$, $x_5^* = 1$; 故原问题的最优解为 $X^* = [1 \ 0 \ 0 \ 0 \ 1]^T$; $w^* = 5$ 。

3. 对偶变量的经济含义——影子价格 (shadow price)

在单纯形法的每步迭代中, 目标函数值 $z = C_B B^{-1} b$ 和检验数 $C_N - C_B B^{-1} N$ 中都有乘子 $Y = C_B B^{-1}$, 那么 Y 的经济含义是什么呢?

设 B 是原问题 $\{\max z = CX \mid AX \leq b, X \geq 0\}$ 的最优基, 则最优值

$$z^* = C_B B^{-1} b = Y^* b$$

因此

$$\frac{\partial z^*}{\partial b} = C_B B^{-1} = Y^*$$

或者说

$$z^* = y_1^* b_1 + y_2^* b_2 + \cdots + y_m^* b_m$$

则

$$\frac{\partial z^*}{\partial b_i} = y_i^*$$

这说明原规划最终表中乘子 $C_B B^{-1}$ 一方面是对偶问题的最优解, 另一方面它还表示原规划中各资源分别增加一个单位时总利润增加多少。这个概念在西方经济学中称为影子价格。如问题 1 中, 终表中松弛变量下检验数

$$\begin{aligned} [\sigma_3 \quad \sigma_4 \quad \sigma_5] &= [0 \quad 0 \quad 0] - C_B B^{-1} [P_3 \quad P_4 \quad P_5] \\ &= -C_B B^{-1} = [0 \quad -1.36 \quad -0.52] \end{aligned}$$

即得乘子 $C_B B^{-1} = [0 \quad 1.36 \quad 0.52]$ 。它们分别称为这个系统中资源煤、电、油的影子价格, 表示煤、电、油各增加一个单位时总利润的增加值。事实上, 当确定最优基 B 后, 目标函数最优值

$$z^* = C_B B^{-1} b = [0 \quad 1.36 \quad 0.52] \begin{bmatrix} 360 \\ 200 \\ 300 \end{bmatrix} = 428$$

以电为例, 资源电由 200 增加到 201 时, 目标函数值

$$\bar{z} = [0 \quad 1.36 \quad 0.52] \begin{bmatrix} 360 \\ 201 \\ 300 \end{bmatrix} = 428 + 1.36$$

利润增加 1.36 个单位。同理, 油的影子价格为 0.52, 煤的影子价格为 0。

显然, 影子价格是约束条件常数项增加一个单位而产生的目标函数最优值的变化量, 这说明对偶变量 y_i 表明了约束条件的影子价格。影子价格是针对某一具体的约束条件而言的, 而问题中所有其他数据都保持不变。这样, 影子价格可以被理解为目标函数最优值对资源的一阶偏导数。另外, 不难发现, 在 $y_i > 0$ 时, b_i 增加后 z^* 将增加; 在 $y_i = 0$ 时, b_i 的增加不会使 z^* 增加。如在问题 1 中, 资源煤的增加不会使利润增加。

下面举例进一步说明影子价格。

例 2.16 某工厂经理对该厂生产的两种产品用线性规划确定最优的产量。根据产品的单位产值和生产这些产品的三种资源供应限量, 建立了如下线性规划模型:

$$\begin{aligned}
 \max z &= 5x_1 + 4x_2 && \text{(产值)} \\
 \begin{cases} x_1 + 3x_2 \leq 90 \\ 2x_1 + x_2 \leq 80 \\ x_1 + x_2 \leq 45 \\ x_1, x_2 \geq 0 \end{cases} &&& \begin{aligned} &\text{(资源 1)} \\ &\text{(资源 2)} \\ &\text{(资源 3)} \end{aligned}
 \end{aligned}$$

对这个线性规划用单纯形法求解,得初始表格和最优表格如表 2.16。

表 2.16

	c_j		5	4	0	0	0	$B^{-1}b$
	C_B	X_B	x_1	x_2	x_3	x_4	x_5	
初始表	0	x_3	1	3	1	0	0	90
	0	x_4	2	1	0	1	0	80
	0	x_5	1	1	0	0	1	45
	σ_j		5	4	0	0	0	$z=0$
最优表	0	x_3	0	0	1	2	-5	25
	5	x_1	1	0	0	1	-1	35
	4	x_2	0	1	0	-1	2	10
	σ_j		0	0	0	-1	-3	$z=215$

这说明最佳生产方案为第一种产品生产 35 件,第二种产品生产 10 件,总产值可以达到 215。

从前面的分析中知道,松弛变量 x_3 、 x_4 和 x_5 的检验数对应着对偶问题的最优解,而这些数值就是这三种资源的影子价格:

资源 1 的影子价格 $y_1 = -\sigma_3 = 0$

资源 2 的影子价格 $y_2 = -\sigma_4 = 1$

资源 3 的影子价格 $y_3 = -\sigma_5 = 3$

因此这三种资源的影子价格分别为 0、1 和 3。

资源 1 的影子价格为零,说明增加这种资源不会增加总的产值。如何来理解这一点呢?可以把上述初始单纯形表中的 90 改为 91,这样最优表将如表 2.17。

表 2.17

c_j		5	4	0	0	0	$B^{-1}b$
C_B	X_B	x_1	x_2	x_3	x_4	x_5	
0	x_3	0	0	1	2	-5	26
5	x_1	1	0	0	1	-1	35
4	x_2	0	1	0	-1	2	10
σ_j		0	0	0	-1	-3	$z=215$

这说明资源 1 的增加不改变产品方案也不增加总的产值。如果资源 2 增加一个单位,从 80 改为 81,最优单纯形表如表 2.18。

表 2.18

c_j		5	4	0	0	0	$B^{-1}b$
0	x_3	0	0	1	2	-5	27.5
5	x_1	1	0	0	1	-1	36
4	x_2	0	1	0	-1	2	9
σ_j		0	0	0	-1	-3	$z = 216$

这说明增加一个单位的资源 2 以后,最佳生产方案为第一种产品生产 36 件,第二种产品生产 9 件,总产值由原来的 215 增加到 216,也就是总产值增加量为 1。当然,有了影子价格的理论就没有必要进行上述计算了。

如果资源 1 和 2 没有变化,而资源 3 增加一个单位,从影子价格 $y_3 = -\sigma_3 = 3$ 可知,总产值的增加量将为 3。值得注意的是,产品的品种没有改变,但是每种产品的生产数量却改变了。容易验证,如果资源 3 增加一个单位,新的规划将是第一种产品生产 34 件,第二种产品生产 12 件,总产值为 218。

影子价格说明了不同资源对总的经济效益产生的影响。因此,一般说来,影子价格对企业的经营管理能够提供一些有价值的信息。

在线性规划应用到经济问题中时,对原始规划可以作这样的解释:变量可以理解为经济活动的水平,如产量,每个可行解就表示一组可行的生产水平。目标函数可以理解为总的经济收益,系数 C 表示每种产品的售价,右端常数项 b 可理解为使用资源的上限,而矩阵 A 中的系数可理解为不同产品对各种资源的单位消耗,对这个线性规划求最优解就是在有限资源条件下谋求最高的收益。此时相应的对偶规划中的最优解就是影子价格。由于影子价格是指资源增加时对最优收益发生的影响,所以有人把它称为资源的边际产出或者资源的机会成本。它表示资源在最优产品组合时,能具有的“潜在价值”或“贡献”。

影子价格在经营管理中的用处很多。一般说来,它可以提供以下几个方面的消息,这里仍以例 2.16 中的数字说明。

①影子价格可以告诉决策者,增加哪一种资源对增加经济效益最有利。如本例中的三种资源影子价格为 0,1,3,说明首先应考虑增加第三种资源,因为相比之下它能使收益增加最多。

②影子价格可以告诉决策者,用多大的代价增加资源才是合算的,如第三种资源每增加一个单位能使收益增加 3,如果增加这种资源的代价大于 3 就不划算了。

③告诉决策者应如何考虑新产品的价格。企业要生产一种新产品时,如果每件新产品耗用的这三种资源数量是 1,2,3 单位,则新产品的定价一定要大于

$$[0 \quad 1 \quad 3] \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 11$$

才能增加公司的收益,如售价低于 11 的话,生产是不划算的。

④使决策者知道产品价格变动时哪些资源最为可贵,哪些无关紧要。如上述例子中,产品的售价不是 [5 4] 而是 [5 5],则从单纯形表格中可以算出影子价格将从 [0 1 3] 改变为

$$[0 \quad 5 \quad 5] \begin{bmatrix} 1 & 2 & -5 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix} = [0 \quad 0 \quad 5]$$

这说明如第二种产品增加价格的话,资源 3 将显得更“宝贵”了。

⑤可以帮助分析工艺改变后对资源节约的收益。如工艺过程改进后使第三种资源能节约 2%,则带来的经济收益将是 $3 \times 45 \times 2\% = 2.7$ 。

以上的分析是有前提的,即最优解的最优基没有变化,具体的分析还要结合下一节的灵敏度分析来进行。

影子价格虽然被定名为一种价格,但是还应对它有更为广义的理解。影子价格是针对约束条件而言的,并不是所有的约束条件都代表了对资源的约束。例如上述规划中还可以列入一个产量约束:两种产品的数量不超过市场上的需要量。这样的约束也有个影子价格。如果这个影子价格算出来比前面几种影子价格都要高的话,则决策者从中得到的信息应理解为:扩大销售量能比增加资源带来更大的经济效益。

2.3.2 灵敏度分析

线性规划方法的应用可分成三步:第一步对实际问题进行细致分析,建立合适的线性规划模型;第二步求解这线性规划模型;第三步对已求出的最优解进行分析。对管理决策者来讲,在求出最优解后,继续进行第三步工作是极为重要的,因为在求解线性规划时都把系数 A 、 b 、 C 作为确定元素考虑。实际上,这些系数往往是变化的,例如市场条件变化会使价格系数 C 改变,技术革新会改变系数矩阵 A 。所谓灵敏度分析工作就是分析系数 A 、 b 、 C 的变化对已得到的最优解有何影响。这里只讨论几种简单情况。

1. 资源向量 b 的分量 b_i 变化的分析

主要分析 b_i 在什么范围内变化不影响最优基。以问题 1 为例,考虑电资源 b_2 在何范围内变化时不影响最优基? 设 b_2 变为 $b_2 + \Delta$,有

$$b = \begin{bmatrix} 360 \\ 200 \\ 300 \end{bmatrix} \rightarrow \bar{b} = \begin{bmatrix} 360 \\ 200 + \Delta \\ 300 \end{bmatrix}$$

由于 b 值变化并不改变检验数 $\sigma_j = c_j - C_B B^{-1} P_j$ 的值,只可能破坏非负条件,因此只要 b 变成 \bar{b} 后,仍然满足非负条件,即有

$$B^{-1} \bar{b} \geq 0$$

则原有的最优基 B 仍是 b 变化后条件下的最优基。从问题 1 终表中知,松弛变量 x_3 、 x_4 、 x_5 下元素列为

$$[P'_3 \quad P'_4 \quad P'_5] = B_2^{-1} [P_3 \quad P_4 \quad P_5] = B_2^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = B_2^{-1}$$

即知最优基 B_2 的逆为

$$B_2^{-1} = [P'_3 \quad P'_4 \quad P'_5] = \begin{bmatrix} 1 & -3.12 & -1.16 \\ 0 & 0.40 & -0.2 \\ 0 & -0.12 & 0.16 \end{bmatrix}$$

为保持非负约束 $B^{-1}b \geq 0$, 有

$$\begin{bmatrix} 1 & -3.12 & 1.16 \\ 0 & 0.40 & -0.20 \\ 0 & -0.12 & 0.16 \end{bmatrix} \begin{bmatrix} 360 \\ 200 + \Delta \\ 300 \end{bmatrix} \geq 0$$

即为

$$\begin{cases} 84 - 3.12\Delta \geq 0 \\ 20 + 0.40\Delta \geq 0 \\ 24 - 0.12\Delta \geq 0 \end{cases}$$

由此得, $-50 \leq \Delta \leq 27$ 或 $150 \leq b_2 \leq 227$ 。

这样就得知, 当电资源在 150 至 227 之间变化时, 原来的方案中, 基 $B_2 = [P_3 \ P_1 \ P_2]$ 仍是最优基。

2. 价格系数 c_j 变化的分析

根据 c_j 对应的变量 x_j 在终表中是基变量还是非基变量, 分别讨论它们的变化对解的影响。

先考虑 c_j 对应基变量的情况。仍考虑问题 1, 目标函数 $z = 7x_1 + 12x_2$ 。现设 c_1 由 7 变化至 $7 + \Delta$, c_2 不变。由于 c_1 是基变量 x_1 之系数, c_1 的变化将影响各检验数, 然而并不破坏非负条件 $B_2^{-1}b \geq 0$ 。

考虑变化后非基变量检验数, 有

$$\begin{aligned} \bar{\sigma}_4 &= c_4 - C_{B_2} B_2^{-1} P_4 = 0 - [0 \ 7 + \Delta \ 12] \begin{bmatrix} -3.12 \\ 0.4 \\ -0.12 \end{bmatrix} = -1.36 - 0.4\Delta \\ \bar{\sigma}_5 &= c_5 - C_{B_2} B_2^{-1} P_5 = 0 - [0 \ 7 + \Delta \ 12] \begin{bmatrix} -1.16 \\ -0.20 \\ 0.16 \end{bmatrix} = -0.52 + 0.20\Delta \end{aligned}$$

为保证原解仍是最优解, 必使 $\bar{\sigma}_4 \leq 0, \bar{\sigma}_5 \leq 0$, 即

$$\begin{cases} -1.36 - 0.4\Delta \leq 0 \\ -0.52 + 0.20\Delta \leq 0 \end{cases}$$

由此得

$$\begin{aligned} -3.4 &\leq \Delta \leq 2.6 \\ 3.6 &\leq c_1 \leq 9.6 \end{aligned}$$

在这个范围内, 最优解不变。

如果考虑 c_j 是对应非基变量的系数, 则问题就简单得多。它的变化只影响它所对应的检验数 σ_j 。因此, 当 c_j 变为 $c_j + \Delta$ 时, 只需保持

$$\bar{\sigma} = c_j + \Delta - C_B B^{-1} P_j = \sigma_j + \Delta \leq 0$$

即可确定不影响最优解之 $\Delta \leq -\sigma_j$ 。

3. 追加新变量的分析

在问题已求出最优解后又追加新的变量(给定相应的价格系数和技术系数)时, 分析最优解的变化。

仍以问题 1 为例, 在问题 1 解完后, 考虑增加新产品丙或丁, 有关数据如表 2.19。问丙产

品和丁产品是否该投产。

表 2.19

	丙	丁
煤(t)	3	12
电(kW·h)	6	3
油(t)	8	4
价格(万元)	10	8

增加丙产品即增加 x_6 变量, $c_6 = 10$, $P_6 = [3 \ 6 \ 8]^T$ 。原终表中基为 $B_2 = [P_3 \ P_1 \ P_2]$ 。它是否仍是最优基, 取决于 x_6 的检验数 σ_6 。而

$$\begin{aligned}\sigma_6 &= c_6 - C_{B_2} B_2^{-1} P_6 \\ &= 10 - [0 \ 7 \ 12] \begin{bmatrix} 1 & -3.12 & 1.16 \\ 0 & 0.4 & -0.20 \\ 0 & -0.12 & 0.16 \end{bmatrix} \begin{bmatrix} 3 \\ 6 \\ 8 \end{bmatrix} \\ &= 10 - [0 \ 1.36 \ 0.52] \begin{bmatrix} 3 \\ 6 \\ 8 \end{bmatrix} = 10 - 12.32 = -2.32 < 0\end{aligned}$$

这说明增加丙产品后, 原最优基 $B_2 = [P_3 \ P_1 \ P_2]$ 仍是最优基, 即 $x_1 = 20$, $x_2 = 24$, $x_3 = 84$, $x_4 = x_5 = x_6 = 0$ 为最优解, 原生产方案仍是最优, 丙产品不应生产。

考虑增加丁产品即增加 x_7 变量, $c_7 = 8$, $P_7 = [12 \ 3 \ 4]^T$ 。同样应计算 x_7 之检验数 σ_7 。

$$\sigma_7 = c_7 - C_{B_2} B_2^{-1} P_7 = 8 - [0 \ 1.36 \ 0.52] \begin{bmatrix} 12 \\ 3 \\ 4 \end{bmatrix} = 8 - 6.16 = 1.84 > 0$$

则原最优基 B_2 已不是最优, 按单纯形法则应取 x_7 为进基变量, 即说明丁产品应投产。它投产会比原方案的收益更大。

这个问题也可以这样分析, 最终表中乘子 $C_B B^{-1} = [0 \ 1.36 \ 0.52]$, 在经济上表示资源煤、电、油的影子价格。这也就是三种资源在系统内部的价格。煤 $y_1 = 0$, 电 $y_2 = 1.36$, 油 $y_3 = 0.52$ 。对新产品丙和丁来说, 每生产单位丙产品需耗煤 3、电 6、油 8。按系统内部价格, 生产单位丙产品的成本为

$$(3 \times 0) + (6 \times 1.36) + (8 \times 0.52) = 12.32$$

而生产单位丙产品收入是 10, 显然每生产 1 个单位丙产品工厂就得亏损 2.32, 所以不应生产丙产品。

同样, 丁产品单位成本为

$$(12 \times 0) + (3 \times 1.36) + (4 \times 0.52) = 6.16$$

而丁产品每单位收入为 8, 说明每生产单位丁产品将盈利 1.84, 所以生产丁产品是有利的。

4. 约束矩阵 A 变化的分析

有下列两种基本情形。

(1) 非基向量列 P_j 改变为 P'_j ,

这种情况指初始表中的 P_j 列数据改变为 P'_j , 而第 j 个列向量在原最终表上是非基向量。

这一改变直接影响最终表上的第 j 列数据和第 j 个检验数。最终表上的第 j 列数据变为 $B^{-1}P_j$, 而新的检验数 $\sigma'_j = c_j - C_B B^{-1}P'_j$ 。

若 $\sigma'_j < 0$, 则原最优解仍是新问题的最优解。若 $\sigma'_j > 0$, 则原最优基在非退化情况下不再是最优基。这时, 应在原来最终表的基础上, 换上改变后的第 j 列数据 $B^{-1}P'_j$ 和 σ'_j , 把 x_j 作为换入变量, 用单纯形法继续迭代。

(2) 基向量列 P_j 改变为 P'_j

这种情况指初始表中的 P_j 列数据改变为 P'_j , 而第 j 个列向量在原最终表上是基向量。此时, 原最优解的可行性和最优性都可能遭到破坏, 问题变得相当复杂, 故一般不去修改原来的最终表, 而是重新计算。

2.4 运输问题

2.4.1 运输问题的一般模型

运输问题就是研究如何制定最合理的物资调运方案, 使总运输费用最低, 运输问题模型不仅适用于物资运输, 也同样适用于其他类似问题。

运输问题的一般模型可由图 2.6 表示。其中有 m 个生产地点(产地)可以供应物资, 用 $A_i (i=1, 2, \dots, m)$ 表示。有 n 个消费地点(销地)需要物资, 用 $B_j (j=1, 2, \dots, n)$ 表示。又知产地 A_i 的产量为 a_i , 销地 B_j 的销量(需求量)为 b_j , 从 A_i 到 B_j 的单位运价为 c_{ij} 。

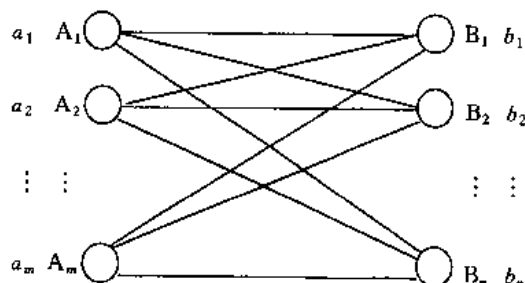


图 2.6

在运输问题模型中, 总产量 $\sum_{i=1}^m a_i$ 等于总销量 $\sum_{j=1}^n b_j$ 时, 称为产销平衡问题。实际工作中, 常用产销平衡表(表 2.20)和单位运价表(表 2.21)描述这类问题。

表 2.20

销地 \ 产地	B_1	B_2	...	B_n	产量
A_1	x_{11}	x_{12}	...	x_{1n}	a_1
A_2	x_{21}	x_{22}	...	x_{2n}	a_2
\vdots	\vdots				\vdots
A_m	x_{m1}	x_{m2}	...	x_{mn}	a_m
销量	b_1	b_2		b_n	$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$

表 2.21

销地	B ₁	B ₂	...	B _n
产地				
A ₁	c ₁₁	c ₁₂	...	c _{1n}
A ₂	c ₂₁	c ₂₂	...	c _{2n}
⋮	⋮			⋮
A _m	c _{m1}	c _{m2}	...	c _{mn}

用 x_{ij} 表示从 A_i 运往 B_j 的运输量。在产销平衡条件下, 运输问题的数学模型为

$$\begin{aligned} \min z = & \sum_{j=1}^n \sum_{i=1}^m c_{ij} x_{ij} \\ \begin{cases} \sum_{j=1}^n x_{ij} = a_i & (i=1, 2, \dots, m) \\ \sum_{i=1}^m x_{ij} = b_j & (j=1, 2, \dots, n) \\ x_{ij} \geq 0 & (i=1, \dots, m, j=1, \dots, n) \end{cases} \end{aligned} \quad (2.8)$$

其中 $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$

显然这是一个线性规划问题, 可以用单纯形法求解。由于这个线性规划自身结构的特殊性, 它的系数矩阵中每列只有两个元素为 1, 其余均为零。因此, 可以用更方便的特殊方法处理。下面介绍的表上作业法就是解决运输问题的常用方法。

2.4.2 表上作业法

表上作业法类似于单纯形法, 也分为三步进行:

第一步, 确定一个初始可行调运方案, 常用的方法有最小元素法、西北角法等;

第二步, 判别当前可行方案是否最优, 常用方法有闭回路法和位势法, 用这两种方法计算出检验数, 从而判别方案是否最优;

第三步, 方案调整, 即从当前方案出发去寻找另一个更好的调运方案。

下面通过例 2.17 说明运输问题表上作业法的具体步骤。

例 2.17 某公司设有三个加工厂 A_1 、 A_2 、 A_3 和四个门市部 B_1 、 B_2 、 B_3 、 B_4 。各加工厂的产量和各门市部的销量如表 2.22 所示。从各加工厂到各门市部的单位运价如表 2.23 所示。在满足各门市部销售需求的情况下, 如何进行调运使总的运输支出最少?

表 2.22

(单位: t)

门市部	B ₁	B ₂	B ₃	B ₄	产量
加工厂					
A ₁					7
A ₂					4
A ₃					9
销量	3	6	5	6	20

表 2.23

(单位:元/t)

加工厂 \ 门市部	B ₁	B ₂	B ₃	B ₄
A ₁	3	11	3	10
A ₂	1	9	2	8
A ₃	7	4	10	5

解 第一步,用最小元素法确定初始调运方案。最小元素法的基本思想就是:就近尽量满足供应,从单位运价表上运价最低的元素确定运输的供应关系,依次类推一直到给出一个初始方案为止,具体步骤如表 2.24 所示。

①在单位运价表上找出最小元素 $c_{21}=1$ 。首先由 A_2 尽量满足 B_1 门市部的需要,最大可能供应量 $x_{21}=3$,在方案表 x_{21} 位置上填写上 3,表示由 A_2 运至 B_1 处 3 个单位。这时 B_1 的需求量已满足,在运价表中划去 B_1 列,表示不必再考虑 B_1 门市部。

②在划去 B_1 列的运价表中,寻找最小元素 $c_{23}=2$,即由 A_2 剩余产量首先尽量供应 B_3 门市部的需求。 B_3 需要 5 个单位, A_2 最大可供 1 个单位。即 $x_{23}=1$ 。在方案表中 x_{23} 位置上填上 1,表示由 A_2 运至 B_3 处 1 个单位, A_2 产量已全部运出,在单位运价表中划去 A_2 行,表示不必再考虑 A_2 了。

③在运价表中没有被直线划去的元素中寻找最小元素 $c_{13}=3$,考虑由 A_1 供应 B_3 处需求, B_3 还需 4 个单位, A_1 能供应 B_3 处 4 个单位,即 $x_{13}=4$,在方案表上 x_{13} 位置上填上 4,表示由 A_1 运至 B_3 处 4 个单位。这时 B_3 处需求已满足,在运价表上划去 B_3 列元素。

④以此类推,继续在运价表上未划去的元素中寻找最小元素 $c_{32}=4$,确定由 A_3 供应 B_2 的最大量 $x_{32}=6$ 。在方案表 x_{32} 位置上填入 6,这时 B_2 处需求已满足,在运价表上划去 B_2 列。

⑤同样,在运价表未划去元素中寻找最小元素 $c_{34}=5$,由 A_3 供应 B_4 的最大可能 $x_{34}=3$,在方案表 x_{34} 位置上填上 3。 A_3 产量已全部运出,在运价表上划去 A_3 行。

⑥继续在运价表中未被划去的元素中寻找最小元素。这时只有一个元素 $c_{14}=10$ 没有被划去,由 A_1 供应 B_4 处,供应量 $x_{14}=3$,在方案表上 x_{14} 位置上填上 3,这时 A_1 产量已全部运出, B_4 的需求量也全部满足,分别在运价表上划去 A_1 行及 B_4 列。此时运价表上所有元素全部被划去。方案表上有数格上数字 $x_{ij} \geq 0$,表示 A_i 往 B_j 处的运量;空格 $x_{ij}=0$,表示 A_i 不往 B_j 处运。表 2.24 中的方案表(一)就构成一个初始方案。

从最小元素法中可知,在方案表上每添一个元素,相应地在运价表上画一直线。运价表上一共画 $m+n$ 条直线,方案表上就相应地有 $m+n-1$ 个有数格。(添加最后一数时,相应画两条直线)很容易证明,由最小元素法得到的初始方案恰好是线性规划中的一基本可行解,空格 $x_{ij}=0$ 对应非基变量,有数格($x_{ij} \geq 0$)对应基变量。另外,当最小元素法某一步确定了最小元素 c_{ij} ,即确定了由 A_i 尽量供应 B_j ,供应量为 x_{ij} 。如果这时恰好 A_i 产量全部运出, B_j 的需求量也正好全部满足,则应在运价表上同时划去 A_i 行和 B_j 列,而在方案表 x_{ij} 位置上填入运量,同时在 i 行或 j 列的任一空格位置上填入 0,以保证有数格有 $m+n-1$ 个。

第二步,判别已确定方案是否最优。

表上作业法是通过计算方案表中空格的检验数判断当前运输方案是否是最优的,计算空

表 2.24

运价表					方案表(一)				
	B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4
A_1	3	11	3	10	A_1				7
① A_2	①	9	2	8	A_2	3			4
A_3	7	4	10	5	A_3				9
						3	6	5	6
	B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4
A_1	3	11	3	10	A_1				7
② A_2	①	9	②	8	A_2	3		1	4
A_3	7	4	10	5	A_3				9
						3	6	5	6
	B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4
A_1	3	11	③	10	A_1			4	7
③ A_2	①	9	②	8	A_2	3		1	4
A_3	7	4	10	5	A_3				9
						3	6	5	6
	B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4
A_1	3	11	③	10	A_1			4	7
④ A_2	①	9	②	8	A_2	3		1	4
A_3	7	④	10	5	A_3			6	9
						3	6	5	6
	B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4
A_1	3	11	②	10	A_1			4	7
⑤ A_2	①	9	②	8	A_2	3		1	4
A_3	7	④	10	⑤	A_3			6	3
						3	6	5	6
	B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4
A_1	3	11	③	⑥	A_1			4	3
⑥ A_2	①	9	②	8	A_2	3		1	4
A_3	7	④	10	⑤	A_3			6	3
						3	6	5	6

格检验数有二种方法,即闭回路法和位势法。

1. 闭回路法

对方案表中每一空格,确定一条由空格出发的闭回路。闭回路是由水平或垂直线组成的闭合图形。闭回路上的顶点除了这个空格外,其余均为有数格。例如,在上述表 2.24 确定的方案表(一)中,空格 x_{11} 出发的闭回路为

$$x_{11} - x_{13} - x_{23} - x_{21} - x_{11}$$

其中除 x_{11} 为空格外,其余均为有数格(表 2.25)。

可以证明,对每一个空格都存在而且惟一存在这样一条封闭回路。在表 2.25 确定的方案表(一)中各空格的封闭回路为

$$x_{11} - x_{13} - x_{23} - x_{21} - x_{11}$$

$$x_{12} - x_{14} - x_{34} - x_{32} - x_{12}$$

$$x_{22} - x_{23} - x_{13} - x_{14} - x_{34} - x_{32} - x_{22}$$

$$x_{24} - x_{23} - x_{13} - x_{14} - x_{24}$$

$$x_{31} - x_{34} - x_{14} - x_{13} - x_{23} - x_{21} - x_{31}$$

$$x_{33} - x_{34} - x_{14} - x_{13} - x_{33}$$

表 2.25

	B ₁	B ₂	B ₃	B ₄
A ₁			4	3
A ₂	3		1	
A ₃		6		3

由各封闭回路可计算各空格的检验数 σ_{ij} 。它等于其闭回路上奇数顶点运价与偶数顶点运价之负值的和,在本例中有

$$\sigma_{11} = c_{11} - c_{13} + c_{23} - c_{21} = 1$$

$$\sigma_{12} = c_{12} - c_{14} + c_{34} - c_{32} = 2$$

$$\sigma_{22} = c_{22} - c_{23} + c_{13} - c_{14} + c_{34} - c_{32} = 1$$

$$\sigma_{24} = c_{24} - c_{23} + c_{13} - c_{14} = -1$$

$$\sigma_{31} = c_{31} - c_{34} + c_{14} - c_{13} + c_{23} - c_{21} = 10$$

$$\sigma_{33} = c_{33} - c_{34} + c_{14} - c_{13} = 12$$

式中 c_{ij} 为封闭回路顶点 x_{ij} 位置上的单位运价。这样得到了方案(一)的检验数表(一)(表 2.26)。

当所有空格检验数 $\sigma_{ij} \geq 0$ 时,则当前方案是最优的。若尚有某空格检验数 $\sigma_{ij} < 0$,则当前方案尚可进一步调整。如检验数表(一)中 $\sigma_{24} = -1$,则方案(一)不是最优方案。

检验数 σ_{ij} 有确切的经济意义。以 σ_{11} 为例,空格 (x_{11}) 表明方案(一)中 A_1 不往 B_1 处运送。现设由 A_1 往 B_1 调运 1 个单位时,观察总运输费用的变化。当 $x_{11} = 0$ 变为 $x_{11} = 1$ 时,为满足约束条件,必须在 (x_{13}) 处减少运送 1 个单位, (x_{23}) 处增运 1 个单位, (x_{21}) 处减运 1 个单位,这时总成本的变化量为

$$c_{11} - c_{13} + c_{23} - c_{21} = 1$$

即每个空格检验数 σ_{ij} 表示,当由 A_i 往 B_j 增运 1 个单位时引起总运输成本的变化数。当所有 $\sigma_{ij} \geq 0$ 时,表示任何一个空格处调运 1 个单位都将引起运输成本上升。这说明当前方案不可能再改进,即定为最优方案。

表 2.26

	B ₁	B ₂	B ₃	B ₄
A ₁	1	2		
A ₂		1		-1
A ₃	10	12		

2. 位势法

表上作业法得到的调运方案对应于运输问题线性规划模型中一个基本可行解。其中有数格对应基变量,空格对应非基变量。这样,可以通过计算单纯形法中的检验数判别方案是否最优。单纯形法中每个变量 x_{ij} 的检验数

$$\sigma_{ij} = c_{ij} - C_B B^{-1} P_{ij}$$

分析运输问题的线性规划模型见式(2.8)。它有 $m+n$ 个约束条件,则对应于可行基 B 的乘子 $C_B B^{-1}$ 有 $m+n$ 个分量。设

$$C_B B^{-1} = [u_1 \quad u_2 \quad \cdots \quad u_m \quad v_1 \quad v_2 \quad \cdots \quad v_n]$$

又知,系数矩阵中每一列

$$P_{ij} = [0 \quad \cdots \underset{(i)}{1} \quad \cdots 0 \quad \cdots \underset{(m+j)}{1} \quad \cdots 0]^T$$

基变量(有数格 $x_{ij} \geq 0$)的检验数 $\sigma_{ij} = 0$,则有

$$\sigma_{ij} = c_{ij} - C_B B^{-1} P_{ij}$$

$$= c_{ij} - [u_1 \quad \cdots \quad u_m \quad v_1 \quad \cdots \quad v_n] \begin{bmatrix} 0 \\ \vdots \\ 1 \quad (i) \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 1 \quad (m+j) \\ \vdots \\ 0 \end{bmatrix} = c_{ij} - (u_i + v_j) = 0$$

即得方程组(对所有有数格 $x_{ij} \geq 0$)

$$u_i + v_j = c_{ij} \quad (2.9)$$

非基变量(空格 $x_{ij} = 0$)的检验数

$$\sigma_{ij} = c_{ij} - C_B B^{-1} P_{ij} = c_{ij} - (u_i + v_j) \quad (2.10)$$

因此,只要确定 $(u_i + v_j)$ 的值就能确定非基变量的检验数 σ_{ij} 。求解方程组(2.9)确定 u_i 和 v_j 的值,代入式(2.10)便能得到非基变量检验数 σ_{ij} 的值。

仍以例 2.17 中方案表(一)为例说明位势法的计算过程。此处, $m=3, n=4$ 。设 $C_B B^{-1} = [u_1 \quad u_2 \quad u_3 \quad v_1 \quad v_2 \quad v_3 \quad v_4]$, 称 u_i 为 A_i 处位势, v_j 为 B_j 处的位势。对已得的方案表(一)中所有有数格(基变量),有方程

$$u_1 + v_3 = c_{13} = 3$$

$$u_2 + v_3 = c_{23} = 2$$

$$u_2 + v_1 = c_{21} = 1$$

$$u_1 + v_4 = c_{14} = 10$$

$$u_3 + v_4 = c_{34} = 5$$

$$u_3 + v_2 = c_{32} = 4$$

上述方程组有 7 个未知数、6 个方程,可设其中任一个变量为一定值。例如,设 $u_1=0$,就能依次解得 $u_1=0, v_3=3, u_2=-1, v_1=2, v_4=14, u_3=-5, v_2=9$ 。然后代入式(2.10),可算出各空格(非基变量)之检验数

$$\sigma_{11} = c_{11} - (u_1 + v_1) = 1$$

$$\sigma_{24} = c_{24} - (u_2 + v_4) = -1$$

$$\sigma_{31} = c_{31} - (u_3 + v_1) = 10$$

$$\sigma_{33} = c_{33} - (u_3 + v_3) = 12$$

$$\sigma_{12} = c_{12} - (u_1 + v_2) = 2$$

$$\sigma_{22} = c_{22} - (u_2 + v_2) = 1$$

把上述检验数列成表格形式,即得与用闭回路法得到的检验数表(一)(表 2.26)完全一致的检验数表。

用位势法计算空格检验数可以在表上进行,主要步骤如下。

1) 计算位势 u_i 和 v_j 的值 在有数格位置上填入相应的运价 c_{ij} (表 2.27)。设 $u_1=0$,在表上位势 u_1 位置上填入 0,然后根据 $c_{ij} = u_i + v_j$ (有数格),依次在表上求得 u_i 和 v_j 值,并填在相应位置上。

2) 计算 $(u_i + v_j)$ 表 把 $(u_i + v_j)$ 位势和值填在表中相应 (x_{ij}) 位置上,并将有数格位置上的 $u_i + v_j$ 值加上括号以示区别,即构成 $(u_i + v_j)$ 表(表 2.28)。

表 2.27

	B_1	B_2	B_3	B_4		
A_1			3	10	u_1	0
A_2	1		2		u_2	-1
A_3		4		5	u_3	-5
	v_1	v_2	v_3	v_4		
	2	9	3	10		

表 2.28

	B_1	B_2	B_3	B_4		
A_1	2	9	(3)	(10)	u_1	0
A_2	1	8	(2)	9	u_2	-1
A_3	-3	(4)	-2	(5)	u_3	-5
	v_1	v_2	v_3	v_4		
	2	9	3	10		

3) 计算检验数表 用单位运价表上元素 c_{ij} 减去 $(u_i + v_j)$ 表上相应位置上元素,即得到检验数表(表 2.29),其中

$$\sigma_{ij} = c_{ij} - (u_i + v_j)$$

表中加括号位置是有数格,不必计算。它的检验数总是 0。

表 2.29

单位运价表(c_{ij})	$(u_i + v_j)$ 表	检验数表(一)(σ_{ij})																																				
<table> <tr><td>3</td><td>11</td><td>3</td><td>10</td></tr> <tr><td>1</td><td>9</td><td>2</td><td>8</td></tr> <tr><td>7</td><td>4</td><td>10</td><td>5</td></tr> </table>	3	11	3	10	1	9	2	8	7	4	10	5	<table> <tr><td>2</td><td>9</td><td>(3)</td><td>(10)</td></tr> <tr><td>(1)</td><td>8</td><td>(2)</td><td>9</td></tr> <tr><td>-3</td><td>(4)</td><td>-2</td><td>(5)</td></tr> </table>	2	9	(3)	(10)	(1)	8	(2)	9	-3	(4)	-2	(5)	<table> <tr><td>1</td><td>2</td><td></td><td></td></tr> <tr><td></td><td>1</td><td></td><td>-1</td></tr> <tr><td>10</td><td></td><td></td><td>12</td></tr> </table>	1	2				1		-1	10			12
3	11	3	10																																			
1	9	2	8																																			
7	4	10	5																																			
2	9	(3)	(10)																																			
(1)	8	(2)	9																																			
-3	(4)	-2	(5)																																			
1	2																																					
	1		-1																																			
10			12																																			

第三步,调整方案。若在检验数表上有某空格的检验数为负,则可改进方案,降低成本。调整的方法是从具有负检验数的空格出发(有多个负检验数时,选择绝对值大的一个),沿它的闭回路进行调整,即在保持方案可行的条件下,尽量增加空格上的运量。以上例中方案(一)的调整予以说明,检验数表(一)中 $\sigma_{24} = -1$,由空格 (x_{24}) 出发的闭回路为(表 2.30)

$$x_{24} - x_{14} - x_{13} - x_{23} - x_{24}$$

从表上易知,空格 (x_{24}) 上增运 θ 个单位,必须在闭回路顶点 (x_{14}) 上减少 θ 个单位, (x_{13}) 上增加 θ 个单位, (x_{23}) 上减少 θ 个单位,才能保持平衡,构成一个可行解。然而,由于 $x_{23}=1, x_{14}=3$,则空格 (x_{24}) 上最多能增加 $\theta=1$ 个单位。一般来说,调整量为闭回路上偶数顶点格上运量的最小数。这样即可调整得可行方案表(二)(表 2.31)。

表 2.30

	B ₁	B ₂	B ₃	B ₄	
A ₁			$(+\theta)4$	$3(-\theta)$	7
A ₂	3		$(-\theta)1$	$(+\theta)$	4
A ₃		6		3	9
	3	6	5	6	

表 2.31

	B ₁	B ₂	B ₃	B ₄	
A ₁			5	2	7
A ₂	3			1	4
A ₃		6		3	9
	3	6	5	6	

方案(二)是否最优?重复第二步计算相应检验数,可得检验数表(二),如表 2.32 所示。表上所有检验数为非负,则方案(二)为最优方案。

表 2.32

	B ₁	B ₂	B ₃	B ₄
A ₁	0	2		
A ₂		2	1	
A ₃	9		12	

需要指出,在调整时,若闭回路上偶数顶点格上出现几个相同的最小数时,这样一次调整会出现两个以上的空格。此时只允许有一个空格,其余出现的空格填上 0,作为有数格看待。例如表

2.33 表示某一运输问题的一个调运方案,若沿空格 (x_{11}) 出发的闭回路

$$\begin{array}{ccccccc} x_{11} & - & x_{31} & - & x_{33} & - & x_{23} & - & x_{12} & - & x_{11} \\ (+\theta) & & (-\theta) & & (+\theta) & & (-\theta) & & (+\theta) & & (-\theta) \end{array}$$

进行调整,调整量 $\theta=2$ 。调整时,空格 (x_{11}) 处增运 $\theta=2$,即 $x_{11}=2$ 。而在 (x_{31}) 、 (x_{23}) 、 (x_{12}) 位置上减少 $\theta=2$,则 $x_{31}=0$ 、 $x_{23}=0$ 、 $x_{12}=0$ 。这时必须在这三个位置上任选一个作为空格,其余二个位置上填入 0(表 2.24),得一个新调运方案。

表 2.33

	B ₁	B ₂	B ₃
A ₁		$(+\theta)$	$2(-\theta)$
A ₂			$5(+\theta)$
A ₃	$2(-\theta)$		$4(+\theta)$

表 2.34

	B ₁	B ₃	B ₃
A ₁	2	0	
A ₂		7	0
A ₃			6

2.4.3 产销不平衡问题

运输问题中总产量和总销量不平衡时,可以增加虚设的产地或销地,把不平衡问题化为产销平衡问题处理。当总产量大于总销量时,即 $\sum_{i=1}^m a_i > \sum_{j=1}^n b_j$ 时,可以增加一虚设销地 B_{n+1} ,其销量设为

$$b_{n+1} = \sum_{i=1}^m a_i - \sum_{j=1}^n b_j$$

相应 (x_{m+1}) 位置上的值表示由 A_i 运往虚拟销地 B_{n+1} 的量,其销量 $b_{n+1} = \sum_{i=1}^m a_i - \sum_{j=1}^n b_j$,实际就是 A_i 没有运出去,在原地贮存的数量,相应运价 $c_{m+1}=0$ 。这样把一个不平衡问题转化为一个平衡问题。产销平衡表和单位运价表如表 2.35 所示。

表 2.35

	B_1	B_2	\cdots	B_n	B_{n+1}	
A_1	x_{11}	x_{12}	\cdots	x_{1n}	x_{1n+1}	a_1
A_2	x_{21}	x_{22}	\cdots	x_{2n}	x_{2n+1}	a_2
\vdots	\vdots				\vdots	\vdots
A_m	x_{m1}	x_{m2}	\cdots	x_{mn}	x_{mn+1}	a_m
	b_1	b_2	\cdots	b_n	b_{n+1}	$= \sum_{i=1}^m a_i - \sum_{j=1}^n b_j$

	B_1	B_2	\cdots	B_n	B_{n+1}
A_1	c_{11}	c_{12}	\cdots	c_{1n}	0
A_2	c_{21}	c_{22}	\cdots	c_{2n}	0
\vdots	\vdots				\vdots
A_m	c_{m1}	c_{m2}	\cdots	c_{mn}	0

同样,当销量大于产量时,即 $\sum_{i=1}^m a_i < \sum_{j=1}^n b_j$,可增添虚拟的产地 A_{m+1} ,产量为 $a_{m+1} = \sum_{j=1}^n b_j - \sum_{i=1}^m a_i$,而相应格上 x_{m+1j} 表示由 A_{m+1} 运往 B_j 的量,实际上就是 B_j 处不能得到满足的数量,相应单位运价 $c_{m+1j} = 0$ 。这样也转化为一产销平衡问题,产销平衡表和单位运价表由 2.36 所示。

表 2.36

	B_1	B_2	\cdots	B_n	
A_1	x_{11}	x_{12}	\cdots	x_{1n}	a_1
A_2	x_{21}	x_{22}	\cdots	x_{2n}	a_2
\vdots	\vdots			\vdots	\vdots
A_m	x_{m1}	x_{m2}	\cdots	x_{mn}	a_m
A_{m+1}	x_{m+11}	x_{m+12}	\cdots	x_{m+1n}	a_{m+1}
	b_1	b_2	\cdots	b_n	

	B_1	B_2	\cdots	B_n
A_1	c_{11}	c_{12}	\cdots	c_{1n}
A_2	c_{21}	c_{22}	\cdots	c_{2n}
\vdots	\vdots			\vdots
A_m	c_{m1}	c_{m2}	\cdots	c_{mn}
A_{m+1}	0	0	\cdots	0

例 2.18 某采石联合公司负责供应 B_1 、 B_2 、 B_3 三处修路工程所需的石子。预计三个工程所需的石子量为 72、102、41 车,联合公司有三个采石厂 A_1 、 A_2 、 A_3 ,每个采石厂能供应石子数为 76、82、77 车。各采石厂至各工程地点之间每车石子运费由表 2.37 给出,怎样安排调运使总运费最低?

表 2.37

单位运价表(单位:元/车)

	B_1	B_2	B_3
A_1	4	8	8
A_2	16	24	16
A_3	8	16	24

解 采石公司总供应能力 $\sum_{i=1}^3 a_i = 235$ 车,而三个工程的总需求量 $\sum_{j=1}^3 b_j = 215$ 车,这是一个总产量大于总销量的不平衡问题,增添一个虚拟的工程 B_4 ,虚拟的销量为

$$b_4 = \sum_{i=1}^3 a_i - \sum_{j=1}^3 b_j = 235 - 215 = 20 \text{ 车}$$

上述问题则转化为一平衡型运输问题,相应产销平衡表与单位运价表由表 2.38 给出。

表 2.38

产销平衡表					单位运价表(单位:元/车)				
	B ₁	B ₂	B ₃	B ₄		B ₁	B ₂	B ₃	B ₄
A ₁					76	4	8	8	0
A ₂					82	16	24	16	0
A ₃					77	8	16	24	0
	72	102	41	20					

对于表 2.38 给出的平衡问题可按上述表上作业法步骤,逐步求得最优方案。

例 2.19 在例 2.18 中,若预计三工程所需的石子量为 72、102、81 车,这时该如何安排调运使总运费最少?

解 公司总供应能力仍为 235 车,而总需求量 $\sum_{j=1}^3 b_j = 255$ 车,这时成了一个总销量大于总产量的不平衡问题,则增添一个虚拟的产地 A₄。虚拟的产量

$$a_4 = \sum_{j=1}^3 b_j - \sum_{i=1}^3 a_i = 255 - 235 = 20 \text{ 车}$$

问题转化为一个平衡问题,相应产销平衡表及单位运价表由表 2.39 给出,可由表上作业法逐步求得最优方案。

表 2.39

	B ₁	B ₂	B ₃			B ₁	B ₂	B ₃
A ₁				76	A ₁	4	8	8
A ₂				82	A ₂	16	24	16
A ₃				77	A ₃	8	16	24
A ₄				20	A ₄	0	0	0
	72	102	81					

2.5 线性整数规划

前面讨论的线性规划中的决策变量都是连续的。然而,许多实际问题中,决策变量往往只能取整数值,例如人数、次数、个数等的变量,把限制部分决策变量或全部决策变量只能取整数的线性规划称为线性整数规划,简称整数规划(Integer Programming),或简称 IP 问题。线性整数规划的一般形式为

$$\begin{aligned} \max z &= \sum_{j=1}^n c_j x_j \\ \begin{cases} \sum_{j=1}^n a_{ij} x_j = b_i & (i=1, 2, \dots, m) \\ x_j \geq 0 & (j=1, 2, \dots, n) \\ x_j \text{ 为整数(部分或全部)} \end{cases} \end{aligned}$$

整数规划一般可分成三类。

1) 纯整数规划 全部决策变量都限定取整数。

2)混合整数规划 部分决策变量限定取整数,即规划中决策变量一部分是整数变量,另一部分是连续变量。

3)0-1型整数规划 不仅限制决策变量为整数,而且只允许取0和1两个值。

整数规划的解法比线性规划复杂,通常使用分支定界法和割平面法。这些方法的基本思想是把一个整数规划问题转化为一系列线性规划问题,通过求解这一系列线性规划问题而最终得到整数规划的最优解。限于篇幅,本书对整数规划的解法不作详细介绍,有兴趣的读者可参考有关著作,这里只简单介绍几个0-1整数规划的实际建模例子。

对于解0-1整数规划问题的隐枚举法(implicit enumeration),本书同样不再做详细阐述。

2.5.1 厂址选择与固定费用问题

1. 厂址选择问题

某地区有 m 座煤矿, i 号矿每年产量为 a_i t,现有火力发电厂一个,每年需用煤 b_0 t,每年运行的固定费用(包括折旧费,但不包括煤的运费)为 h_0 元。现规划新建一个发电厂, m 座煤矿每年开采的原煤将全部供给这两个电厂发电用。现有几个备选的厂址,若在 j 号备选厂址建电厂,每年运行的固定费用为 h_j ,每吨原煤从 i 号矿运送到 j 号备选厂址的运费为 c_{ij} ($i=1, \dots, m; j=1, \dots, n$),每吨原煤从 i 号矿运送到原有电厂的运费为 c_{i0} ($i=1, 2, \dots, m$)。试问,应把新厂厂址选在何处, m 座煤矿开采的原煤应如何分配给两个电厂,才能使每年总费用(电厂运行的固定费用与原煤运费之和)最小?

这样的选址问题可以用0-1型整数规划求解。从条件可知,新建电厂每年用煤量

$$b = \sum_{i=1}^m a_i - b_0$$

令决策变量

$$y_j = \begin{cases} 1, & j \text{ 号备选址被选中 } (j=1, 2, \dots, n) \\ 0, & \text{否则} \end{cases}$$

为了方便,称原有电厂为0号厂,在 j 号备选厂址处建的厂为 j 号厂。设 x_{ij} 为每年从 i 号矿运送到 j 号厂的原煤数量($i=1, 2, \dots, m; j=0, 1, \dots, n$)。于是,每年总费用等于

$$z = \sum_{i=1}^m \sum_{j=0}^n c_{ij} x_{ij} + \sum_{j=1}^n h_j y_j + h_0$$

若 j 号备选厂址未被选中,即 $y_j=0$,那么 j 厂根本就不存在,这时 x_{ij} 应全为零,故有下列约束条件:

$$\begin{aligned} \sum_{j=0}^n x_{ij} &= a_i \quad (i=1, 2, \dots, m) \\ \sum_{i=1}^m x_{i0} &= b_0 \\ \sum_{i=1}^m x_{ij} &= b y_j \quad (j=1, 2, \dots, n) \\ x_{ij} &\geq 0 \quad (i=1, 2, \dots, m; j=0, 1, 2, \dots, n) \end{aligned}$$

又因只需建一个电厂,故还应有约束条件

$$\sum_{j=1}^n y_j = 1$$

于是,上述选址问题可以归纳成下列形式的整数规划:

$$\begin{aligned} \min z = & \sum_{i=1}^m \sum_{j=0}^n c_{ij} x_{ij} + \sum_{j=1}^n h_j y_j + h_0 \\ \text{s. t. } & \begin{cases} \sum_{j=0}^n x_{ij} = a_i, & (i=1, 2, \dots, m) \\ \sum_{i=1}^m x_{i0} = b_0 \\ \sum_{i=1}^m x_{ij} = b_j y_j, & (j=1, 2, \dots, n) \\ \sum_{j=1}^n y_j = 1 \\ x_{ij} \geq 0 & (i=1, 2, \dots, m, j=0, 1, \dots, n) \\ y_j = 0 \text{ 或 } 1 & (j=1, 2, \dots, n) \end{cases} \end{aligned}$$

2. 固定费用问题

上述的厂址选择问题实际上与下述固定费用问题属同一类问题。一个固定费用问题完整的数学模型用矩阵形式可表示为

$$\begin{aligned} \min z = & KY + CX \\ \text{s. t. } & \begin{cases} AX = b \\ X \leq MY \\ X \geq 0 \\ Y \text{ 为 } 0-1 \text{ 向量} \end{cases} \end{aligned}$$

式中, K 为固定成本向量; C 为变动成本向量; X 为一般的决策向量; Y 为 0-1 决策向量; b 为资源向量; A 为技术系数矩阵; M 的分量为相应于决策变量的上界, 它是为表达当 $x_j > 0$ 时, $y_j = 1$ 这种逻辑关系而设立的, 具体含义在下例中详细说明。

如某工厂为了生产某种产品, 有几种不同的生产方式可供选择, 如选定的生产方式投资高 (选购自动化程度高的设备), 由于产量大, 因而分配到每件产品的变动成本就降低; 反之, 如选定的生产方式投资低, 将来分配到每件产品上的变动成本可能增加, 因此必须全面考虑。现设有三种生产方式可供选择, 其他的资源约束表示为 $AX = b$ 。令 x_j 表示采用第 j 种生产方式时的产量; c_j 表示采用第 j 种生产方式时每件产品的变动成本; k_j 表示采用第 j 种生产方式时的固定成本。于是生产第 j 种产品的总成本函数可写成:

$$P_j(x_j) = \begin{cases} k_j + c_j x_j, & x_j > 0 \quad (j=1, 2, 3) \\ 0, & x_j = 0 \end{cases}$$

另设

$$y_j = \begin{cases} 0, & x_j = 0, \text{ 当不选第 } j \text{ 种生产方式} \quad (j=1, 2, 3) \\ 1, & x_j > 0, \text{ 当选择第 } j \text{ 种生产方式} \end{cases}$$

这样, 初步考虑固定费用问题可用如下模型表示:

$$\min z = (k_1 y_1 + c_1 x_1) + (k_2 y_2 + c_2 x_2) + (k_3 y_3 + c_3 x_3)$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^3 a_{ij}x_j = b_i & (i=1,2,\dots,m) \\ x_j \geq 0 \\ y_j = 0 \text{ 或 } 1 & (j=1,2,3) \end{cases}$$

根据题意,应该有 $x_j > 0, y_j = 1$ 。它的含义是,当生产 x_j 任何正的数量时,都应支付一笔固定费用 $k_j > 0$;而当 $x_j = 0$,则 $y_j = 0$ 。在极小化问题中, $x_j = 0$ 自动蕴涵着 $y_j = 0$ 。因为 $x_j = 0, y_j = 1$ 时,总可以置 $y_j = 0$ 使目标函数值变小,以获得更优的解,这样并未破坏其他约束条件。然而,当 $x_j > 0$ 时, $y_j = 1$ 这种逻辑关系在上述模型形式中并没有得以体现。为此,取 x_j 的一个上界 M_j ,使得 $x_j \leq M_j$,对满足 $AX \leq b, X \geq 0$ 的 x_j 值都成立,并在上述模型中加入约束条件

$$x_1 \leq M_1 y_1, x_2 \leq M_2 y_2, x_3 \leq M_3 y_3$$

则该问题的数学模型可写成

$$\begin{aligned} \min z &= \sum_{i=1}^3 (k_i y_i + c_i x_i) \\ \text{s. t. } &\begin{cases} \sum_{j=1}^3 a_{ij}x_j = b_i & (i=1,2,\dots,m) \\ x_j \geq 0 \\ x_j \leq M_j y_j \\ y_j = 0 \text{ 或 } 1 & (j=1,2,3) \end{cases} \end{aligned}$$

例 2.20 某地区有甲、乙两个城镇,它们每周分别产生 700 t 和 1 200 t 固体废物,现拟用三种方式(焚烧、填海、掩埋)分别在 1、2、3 三个场地对这些废物进行处理(图 2.7),每个处理场的处理费用分为固定费用和可变费用两部分,连同处理场的处理能力分别列于表 2.40 和表 2.41。试求使两城镇处理固体废物总费用最小的方案。

表 2.40

费用和能力	固定费用(元/周)	可变费用(元/t)	处理能力(t/周)
处理场所			
1. 焚烧	3 850	12	1 000
2. 填海	1 150	16	500
3. 掩埋	1 920	6	1 300

解 设 x_{ij} 为由城镇 i ($i=1,2$, 表示甲、乙城镇)运到处理场 j ($j=1,2,3$)处理的废物量。

表 2.41

运费(元/t)	处理场	1	2	3
城 镇				
甲		7.5	2.5	15.0
乙		5.0	7.5	12.5

又设 0-1 变量

$$y_j = \begin{cases} 1 & \text{当第 } j \text{ 个处理场被启用,即 } x_{1j} + x_{2j} > 0 \text{ 时} \\ 0 & \text{当第 } j \text{ 个处理场不被启用,即 } x_{1j} + x_{2j} = 0 \text{ 时} \end{cases}$$

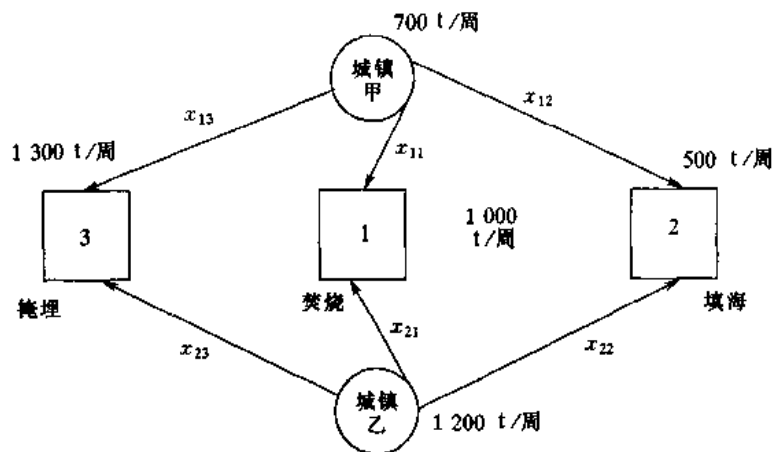


图 2.7

这样,第一个处理场的固定费用可表示为 $3850y_1$;第二个处理场的固定费用可表示为 $1150y_2$;第三个处理场的固定费用可表示为 $1920y_3$ 。这相当于 KY 。 $K = [3850 \ 1150 \ 1920]$, $Y = [y_1 \ y_2 \ y_3]^T$ 。

由城镇甲运 1 t 废物到处理场 1 进行处理的总可变费用为运费和可变处理费用之和,即 x_{11} 的可变费用系数为 $12 + 7.5 = 19.5$ 元/t;其他依此类推。再加上各处理场的固定费用,则总费用函数为

$$z = 19.5x_{11} + 17.0x_{21} + 18.5x_{12} + 23.5x_{22} + 21.0x_{13} + 18.5x_{23} \\ + 3850y_1 + 1150y_2 + 1920y_3$$

约束条件应包括下述几个方面。

每个城镇所有运出的废物等于它产生的废物:

$$\text{对城镇甲} \quad \sum_{j=1}^3 x_{1j} = 700 \quad (2.9)$$

$$\text{对城镇乙} \quad \sum_{j=1}^3 x_{2j} = 1200 \quad (2.10)$$

对各处理场,每个处理场可处理的废物总量不应超过其处理能力:

$$\text{对处理场 1} \quad \sum_{i=1}^2 x_{i1} \leq 1000$$

$$\text{对处理场 2} \quad \sum_{i=1}^2 x_{i2} \leq 500$$

$$\text{对处理场 3} \quad \sum_{i=1}^2 x_{i3} \leq 1300$$

当然还应有约束, $x_{ij} \geq 0 (i=1,2, j=1,2,3), y_j = 0$ 或 1。

应该注意,如果只有这些约束条件,再加上目标函数求极小化,那么极小化的结果必然导致所有的 $y_j = 0 (j=1,2,3)$,而由于有约束条件式(2.9)、式(2.10),不可能所有的 $x_{ij} = 0$ 。这就不能保证 x_{ij} 与 y_j 之间的“逻辑”关系。为解决这个问题,必须再加入三个约束条件:

$$x_{1j} + x_{2j} \leq M_j y_j \quad (j=1,2,3) \quad (2.11)$$

这里 M_j 是一个充分大的正数,至少要大于 $x_{1j} + x_{2j}$ 可能的最大值。根据三个处理场的最大

处理能力,不难给出一个合适的 M_j 值,如 $M_1 \geq 1\,000$ 即可。

有了约束式(2.11),则可保证当 $x_{1j} + x_{2j} > 0$ 时,必有 $y_j = 1$,则式(2.11)可以和其余三个约束条件合并为以下三式:

$$\sum_{i=1}^2 x_{i1} \leq 1\,000 y_1 \quad \sum_{i=1}^2 x_{i2} \leq 500 y_2 \quad \sum_{i=1}^2 x_{i3} \leq 1\,300 y_3$$

这样完整的模型为:

$$\begin{aligned} \min z = & 19.5x_{11} + 17.0x_{21} + 18.5x_{12} + 23.5x_{22} + 21.0x_{13} + 18.5x_{23} \\ & + 3\,850y_1 + 1\,150y_2 + 1\,920y_3 \end{aligned}$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^3 x_{1j} = 700 \\ \sum_{j=1}^3 x_{2j} = 1\,200 \\ \sum_{i=1}^2 x_{i1} = 1\,000 y_1 \\ \sum_{i=1}^2 x_{i2} \leq 500 y_2 \\ \sum_{i=1}^2 x_{i3} \leq 1\,300 y_3 \\ x_{ij} \geq 0, y_j = 0 \text{ 或 } 1 \quad (i=1,2; j=1,2,3) \end{cases}$$

可得该模型的解为

$$y_1^* = 1, x_{11}^* = 200 \text{ t/周}$$

$$y_2^* = 1, x_{12}^* = 500 \text{ t/周}$$

$$y_3^* = 1, x_{21}^* = 800 \text{ t/周}$$

$$x_{23}^* = 400 \text{ t/周}$$

解的结果如图 2.8。

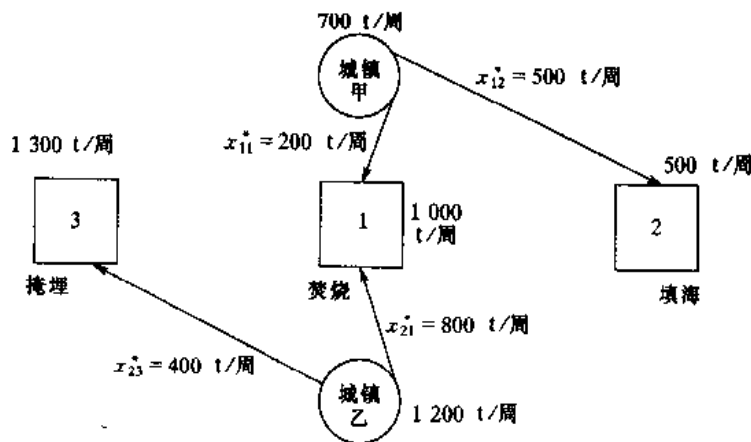


图 2.8

2.5.2 多种类、多方案投资决策问题

在许多复杂的情况下,最优投资方案的选择可归结为整数规划问题。

有一笔投资 p , 对 k 种产品投资, 第 i 种产品有 m_i 个投资方案。第 i 种产品的第 j 个方案需投资为 p_{ij} , 相应每年收益为 M_{ij} ($i=1, 2, \dots, k; j=1, 2, \dots, m_i$)。投资为 n 年, 如资金总额用不完, 则以平均利率 r 作为普通投资(或作其他企业投资), 试作出最优投资决策。

为此, 引入 0-1 型决策变量:

$$x_{ij} = \begin{cases} 1, & \text{表示选用第 } i \text{ 种产品的第 } j \text{ 种方案} \\ 0, & \text{表示不选用第 } i \text{ 种产品的第 } j \text{ 种方案} \end{cases}$$

由于最多只能在 i 产品的 m_i 个方案中选用一个, 并且总资金 $\sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} p_{ij}$ 应不超过 p , 则约束条件为

$$\begin{aligned} \sum_{j=1}^{m_i} x_{ij} &\leq 1 \quad (i=1, 2, \dots, k) \\ \sum_{i=1}^k \sum_{j=1}^{m_i} p_{ij} x_{ij} &\leq p \end{aligned}$$

每年的投资收益有两部分, 一为所用投资资金的收益

$$f_1(\mathbf{X}) = \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} M_{ij}$$

另一为剩余资金的收益, 剩余资金为

$$p - \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} p_{ij}$$

将其转化为普通投资(以平均利率 r 计算), 则剩余资金的收益为

$$f_2(\mathbf{X}) = (p - \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} p_{ij}) \frac{r(1+r)^n}{(1+r)^n - 1}$$

式中, $\frac{r(1+r)^n}{(1+r)^n - 1}$ 为资金回收系数。

目标函数即为每年的总收益

$$z = f_1(\mathbf{X}) + f_2(\mathbf{X}) = \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} M_{ij} + (p - \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} p_{ij}) \frac{r(1+r)^n}{(1+r)^n - 1}$$

整个投资决策归结于求解下列整数规划问题:

$$\begin{aligned} \max z &= \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} M_{ij} + (p - \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} p_{ij}) \frac{r(1+r)^n}{(1+r)^n - 1} \\ \text{s. t. } &\begin{cases} \sum_{j=1}^{m_i} x_{ij} \leq 1 & (i=1, 2, \dots, k) \\ \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} p_{ij} \leq p \\ x_{ij} = 0 \text{ 或 } 1 & (i=1, 2, \dots, k; j=1, 2, \dots, m_i) \end{cases} \end{aligned}$$

例 2.21 设有资金 550 万元,拟对三种产品投资。第一种产品有 2 个投资方案,第二、三种产品各有 1 个投资方案,各方案需投资数及其利润率如表 2.42 所示。剩余资金作为普通投资用,利率 $r=10\%$, $n=5$ 。问该如何作出决策,以使收益最大?

表 2.42

产品方案	第一种产品		第二种产品	第三种产品
	1	2	1	2
投资额 p_{ij} (万元)	300	280	260	240
达到的利润率 \bar{r}	30%	28%	28%	26%

解 据题意可知, $p=550$ 万元, $k=3$, $m_1=2$, $m_2=m_3=1$,决策变量 x_{11} 、 x_{12} 、 x_{21} 、 x_{31} 为 0-1 变量,约束条件为

$$\sum_{j=1}^2 x_{1j} = x_{11} + x_{12} \leq 1$$

$$x_{21} \leq 1$$

$$x_{31} \leq 1$$

$$\sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} p_{ij} = 300x_{11} + 280x_{12} + 260x_{21} + 240x_{31} \leq 550$$

目标函数 $z = f_1(\mathbf{X}) + f_2(\mathbf{X})$, 其中

$$f_1(\mathbf{X}) = \sum_{i=1}^k \sum_{j=1}^{m_i} x_{ij} M_{ij} = M_{11}x_{11} + M_{12}x_{12} + M_{21}x_{21} + M_{31}x_{31}$$

这里 $M_{ij} = p_{ij} \frac{\bar{r}(1+\bar{r})^n}{(1+\bar{r})^n - 1}$, 将相应的参数 \bar{r} 代入, 即得 M_{11} 、 M_{12} 、 M_{21} 、 M_{31} 。而

$$f_2(\mathbf{X}) = [550 - (300x_{11} + 280x_{12} + 260x_{21} + 240x_{31})] \frac{0.1(1+0.1)^5}{(1+0.1)^5 - 1}$$

问题归结为求解整数规划:

$$\max z = f_1(\mathbf{X}) + f_2(\mathbf{X})$$

$$\text{s. t. } \begin{cases} x_{11} + x_{12} \leq 1 \\ 300x_{11} + 280x_{12} + 260x_{21} + 240x_{31} \leq 550 \\ x_{11}, x_{12}, x_{21}, x_{31} = 0 \text{ 或 } 1 \end{cases}$$

求解结果为

$$x_{11} = 1, x_{12} = 0, x_{21} = 0, x_{31} = 1, z = 231$$

2.5.3 运动员的选拔问题

最优化方法在体育事业的各个领域都有着广泛的应用,如在很多情况下,运动员的选拔问题便可以归结为 0-1 整数规划问题。以下以选拔篮球运动员和游泳运动员为例,阐述整数规划在选拔运动员中的作用。

1. 篮球运动员的选拔问题

在长 28 m、宽 15 m 的场地上,两支篮球劲旅龙争虎斗。谁胜谁负,往往取决于多种因素。但是,多数篮球教练却对运动员的身高情有独钟。这是可以理解的,因为篮圈直径只有 45 cm,而且离地而高达 3.05 m,致使投篮的命中离、盖帽和争抢篮板球的成功率无不与篮球运动

员的身高呈正相关。下面,就以身高为主要指标,讨论篮球运动员的选拔问题。

例 2.22 某篮球队拟由编号为 1、2、……、6 的 6 名预备队员中,挑选 3 名正式队员,要求他们的平均身体尽量高。此外,入选队员尚须符合下列条件:①至少有 1 名后卫;②2 号和 5 号至多只许入选 1 名;③最多入选 1 名中锋;④若 2 号或 4 号入选,6 号就不得入选。这些预备队员的有关情况见表 2.43。哪 3 名预备队员应该入选?

表 2.43

预备队员编号	位置	身高(m)
1	中锋	1.93
2	中锋	1.91
3	前锋	1.87
4	前锋	1.86
5	后卫	1.80
6	后卫	1.85

解 设 $x_j = \begin{cases} 0, & \text{表示不选 } j \text{ 号预备队员} \\ 1, & \text{表示选 } j \text{ 号预备队员} \end{cases} \quad (j=1,2,\dots,6), y$ 为 3 名入选者的平均身高。

根据题意,例 2.22 的数学模型为

$$\max y = \frac{1}{3}(1.93x_1 + 1.91x_2 + 1.87x_3 + 1.86x_4 + 1.80x_5 + 1.85x_6)$$

$$\begin{cases} x_1 + x_2 + \dots + x_6 = 3 \\ x_5 + x_6 \geq 1 \\ x_2 + x_5 \leq 1 \\ x_1 + x_2 \leq 1 \\ x_2 + x_6 \leq 1 \\ x_4 + x_6 \leq 1 \\ x_1, x_2, \dots, x_6 = 0 \text{ 或 } 1 \end{cases}$$

求得最优解为 $[x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6] = [1 \ 0 \ 1 \ 0 \ 0 \ 1]$, 最优值为 $y_{\max} = 1.88 \text{ m}$ 。

由上述看到,在表示最优解的 6 维向量中,分量 x_1, x_3, x_6 全等于 1。于是可以断言,1 号、3 号和 6 号预备队员应该入选。最优值表明,这 3 名入选者的平均身高为 1.88 m。

2. 游泳运动员的选拔问题

4×100 m 混合泳接力是观众最感兴趣的游泳项目之一。现在,从一个实例出发,研究混合泳运动员的选拔问题。

例 2.23 甲、乙、丙、丁是 4 名游泳运动员,他们各种姿势的 100 m 游泳成绩见表 2.44。为组成一个 4×100 m 混合泳接力队,怎样选派运动员,方使接力队的游泳成绩最好?

表 2.44

运动员	仰泳	蛙泳	蝶泳	自由泳
甲	75.5	86.8	66.6	58.4
乙	65.8	66.2	57.0	52.8

续表

运动员	仰泳	蛙泳	蝶泳	自由泳
丙	67.6	84.3	77.8	59.1
丁	74.0	69.4	60.8	57.0

解 设 $i=1,2,3,4$ 分别表示甲、乙、丙、丁; $j=1,2,3,4$ 分别表示仰泳、蛙泳、蝶泳、自由泳。设

$$x_{ij} = \begin{cases} 0, & \text{表示 } i \text{ 不参加 } j \\ 1, & \text{表示 } i \text{ 参加 } j \end{cases}$$

y 为混合泳接力队的成绩。

根据题意,例 2.23 的数学模型为

$$\begin{aligned} \min y = & 75.5x_{11} + 86.8x_{12} + 66.6x_{13} + 58.4x_{14} + 65.8x_{21} + 66.2x_{22} + 57.0x_{23} \\ & + 52.8x_{24} + 67.6x_{31} + 84.3x_{32} + 77.8x_{33} + 59.1x_{34} + 74.0x_{41} + 69.4x_{42} \\ & + 60.8x_{43} + 57.0x_{44} \\ \begin{cases} x_{1j} + x_{2j} + x_{3j} + x_{4j} = 1 & (j=1,2,3,4) \\ x_{i1} + x_{i2} + x_{i3} + x_{i4} = 1 & (i=1,2,3,4) \\ x_{ij} = 0 \text{ 或 } 1 & (i,j=1,2,3,4) \end{cases} \end{aligned}$$

这类数学模型称为指派问题,是 0-1 规划的特例,可采用匈牙利法来求解。按照这种解法进行计算,求得最优解为

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

最优值为 $\min y = 252.4 \text{ s}$

不难看出,在表示最优解的四阶矩阵中,元素 x_{14} 、 x_{23} 、 x_{31} 、 x_{42} 皆等于 1。由此可知,选派甲参加自由泳、乙参加蝶泳、丙参加仰泳、丁参加蛙泳,可使混合泳接力队的游泳成绩最好。并且最优值表明,这支援力队只消 252.4 s 便能游完全程。

2.5.4 背包问题

背包问题是最简单的整数规划问题之一。它可解释为一个旅行者,为了准备旅行的必备物品,要在背包里装一些最有用的东西,他最多只能携带 $b \text{ kg}$ 重的物品,而每件物品都只能整件携带,于是他给每件物品规定了一定的“价值”,以表示其有用程度。如果共有 m 件物品可供选择,第 j 件物品重 $a_j \text{ kg}$,其价值为 c_j 。问在携带的物品总重量不超过 $b \text{ kg}$ 的条件下,携带哪些物品,可使总价值最大。背包问题可用如下数学模型表示:

引入 0-1 变量 x_j , 设

$$x_j = \begin{cases} 1 & \text{携带第 } j \text{ 件物品} \\ 0 & \text{不携带第 } j \text{ 件物品} \end{cases}$$

则

$$\begin{aligned} \max z &= \sum_{j=1}^m c_j x_j \\ \text{s. t. } &\begin{cases} \sum_{j=1}^m a_{ij} x_j \leq b \\ x_j = 0 \text{ 或 } 1 \end{cases} \end{aligned}$$

该数学模型仅受限制于单个线性约束条件,变量要求取 0、1,称做“0-1”背包问题。

背包问题去掉整数约束条件变成一普通的线性规划问题是极易求解的。只要将变量按照它们在目标函数中的系数与约束条件中系数之比的大小进行排列。即按“单位重量的价值”的大小进行排列,如

$$\frac{c'_1}{b'_1} \geq \frac{c'_2}{b'_2} \geq \frac{c'_3}{b'_3}$$

其中 1'、2'……是按比值大小排列后的新编号。在不破坏约束条件的情况下,取 x'_1 的值尽可能大,再根据已取定的 x'_1 值,取 x'_2 的值尽可能大,依次类推。这种算法也称做“贪婪算法”但对变量取整数的背包问题,贪婪算法只是一种近似算法。

例 2.24 一个徒步旅行者要在背包中选择一些最有价值的物品携带。他最多能携带 115 kg 的物品。现共有 5 件物品,分别重 54 kg、35 kg、57 kg、46 kg、19 kg,其价值依次为 7、5、9、6、3。问该旅行者携带哪些物品,使总价值最大?

解 设

$$x_j = \begin{cases} 1 & \text{携带第 } j \text{ 件物品} \\ 0 & \text{不携带第 } j \text{ 件物品} \end{cases}$$

则有

$$\begin{aligned} \max z &= 7x_1 + 5x_2 + 9x_3 + 6x_4 + 3x_5 \\ \text{s. t. } &\begin{cases} 54x_1 + 35x_2 + 57x_3 + 46x_4 + 19x_5 \leq 115 \\ x_j = 0 \text{ 或 } 1 \quad (j=1, 2, \dots, 5) \end{cases} \end{aligned}$$

“单位重量价值”由大到小的顺序为

$$\frac{3}{19} = \frac{9}{57} > \frac{5}{35} > \frac{6}{46} > \frac{7}{54}$$

在满足约束条件的情况下,应取 $\frac{3}{19}$ 及 $\frac{9}{57}$ 对应的变量 x_5, x_3 尽可能的大,取 $x_5 = x_3 = 1$ 。

同理根据已取定的 x_5, x_3 的值,取 $\frac{5}{35}$ 对应的变量 x_2 尽可能的大。0-1 背包问题的最优解为

$$x_5 = 1 \quad x_3 = 1 \quad x_2 = 1 \quad x_4 = x_1 = 0$$

类似背包问题的问题还有工厂下料问题及运输货物装载问题等

习 题

2.1 某工厂在计划期内要安排生产 I、II 两种产品,这些产品分别需要在 A、B、C、D 四种不同的设备上加工。按工艺规定,产品 I 和 II 在各设备上所需要的加工台时数及有关数据如表 2.45。现工厂需拟定使总利润最大的生产计划。(1)建立此问题的数学模型;(2)用图解法求解最优计划和最大利润。

表 2.45

所需台时 产品		设备				每件利润(百元)
		A	B	C	D	
I		2	1	4	0	2
II		2	2	0	4	3
台时限制		12	8	16	12	

2.2 分别用图解法和单纯形法求解下面的线性规划,并指出单纯形法迭代的每一步相当于图形上哪一个顶点。

$$\max z = 2x_1 + x_2$$

$$\begin{cases} 3x_1 + 5x_2 \leq 15 \\ 6x_1 + 2x_2 \leq 24 \\ x_1, x_2 \geq 0 \end{cases}$$

2.3 用大 M 法求解下面的线性规划:

$$\max z = 2x_1 + 3x_2 - 5x_3$$

$$\begin{cases} x_1 + x_2 + x_3 = 7 \\ 2x_1 - 5x_2 + x_3 \geq 10 \\ x_1, x_2, x_3 \geq 0 \end{cases}$$

2.4 某昼夜服务的公交线路每天各时间区段内所需司机和乘务人员数如下:

班次	时间	所需人数
1	6:00~10:00	60
2	10:00~14:00	70
3	14:00~18:00	60
4	18:00~22:00	50
5	22:00~ 2:00	20
6	2:00~ 6:00	30

设司机和乘务人员分别在各时间区段一开始时上班,并连续工作 8 小时,问该公交线路至少配备多少名司乘人员。建立该问题的数学模型。

2.5 在解一线性规划问题时,使用单纯形法,已知单纯形表中的初表及终表如表 2.46。

表 2.46

	C_B	X_B	$B^{-1}b$	2	-1	1	0	0	0
				x_1	x_2	x_3	x_4	x_5	x_6
初表	0	x_4	60	3	1	1	1	0	0
	0	x_5	10	1	-1	2	0	1	0
	0	x_6	20	1	1	-1	0	0	1
终表		σ_j		2	-1	1	0	0	0
		x_4						-1	-2
		x_1						1/2	1/2
		x_2						-1/2	1/2
		σ_j							

(1) 写出最优基 B 和 B^{-1} ;

(2) 填完终表;

(3)指出最优解和最优值。

2.6 写出下列线性规划问题的对偶模型:

$$(1) \max z = 5x_1 + 3x_2 + 2x_3 + 4x_4$$

$$\begin{cases} 5x_1 + x_2 + x_3 + 8x_4 = 10 \\ 2x_1 + 4x_2 + 3x_3 - 2x_4 = 10 \\ x_i \geq 0, i = 1, 2, 3, 4 \end{cases}$$

$$\begin{cases} 2x_1 + 4x_2 + 3x_3 - 2x_4 = 10 \\ x_i \geq 0, i = 1, 2, 3, 4 \end{cases}$$

$$x_i \geq 0, i = 1, 2, 3, 4$$

$$(2) \min z = 2x_1 + 2x_2 + 4x_3$$

$$\begin{cases} 2x_1 + 3x_2 + 5x_3 \geq 2 \\ 3x_1 + x_2 + 7x_3 \leq 3 \\ x_1 + 4x_2 + 6x_3 \leq 5 \\ x_1, x_2 \geq 0 \end{cases}$$

$$3x_1 + x_2 + 7x_3 \leq 3$$

$$x_1 + 4x_2 + 6x_3 \leq 5$$

$$x_1, x_2 \geq 0$$

2.7 某电视机厂生产四种型号的特种电视机: I 型——轻便黑白, II 型——正规黑白, III 型——轻便彩色, IV 型——正规彩色。各型号每台所需组装时间、调试时间、销售收入以及该厂组装调试能力如表 2.47 所示。

表 2.47

	I	II	III	IV	厂能力
组装时间(h)	8	10	12	15	2 000
调试时间(h)	2	2	4	5	500
售 价(百元)	4	6	8	10	

但现在显像管紧缺,每月最多只能进货 180 只,其中彩色显像管不超过 100 只,令 x_1, x_2, x_3, x_4 依次表示各型号每月计划产量。现工厂需拟定使目标总销售收入 z 为最大的生产计划。

(1)写出该问题的数学模型,对于约束条件依下列次序:组装时间、调试时间、显像管数、彩色显像管数,并引入松弛变量使之成为等式。

(2)用单纯形法求解得终表如表 2.48。

表 2.48

C_B	X_B	$B^{-1}b$	4	6	8	10	0	0	0	0
			x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0	x_8	50	0.2	0	0.2	0	0.1	-0.5	0	1
6	x_2	125	0.5	1	0	0	0.25	-0.75	0	0
0	x_7	5	0.3	0	0.2	0	-0.15	0.25	1	0
10	x_4	50	0.2	0	0.8	1	-0.1	0.5	0	0
σ_j			-1	0	0	0	-0.5	-0.5	0	0

试分别回答:

(1)最优生产计划是什么? 是否还有其他最优生产计划? 为什么?

(2)组装时间的影子价格是多少?

(3)若外厂可调剂增加 80 h 的调试时间,但每小时需付 0.4(百元),这样的调剂值得吗? 能增加多少收入?

(4)设 I 型机售价由 4(百元)增加到 4.5(百元),最优计划要改变吗? 如果增加到 5.5(百元)呢? 说明理由。

(5)写出本问题的对偶模型,并指出其最优解。

2.8 某照相机厂生产 A、B、C 三种型号的产品,每架均需经机身制造、零件装配和检验包装三道工序。各型号产品所需工作时数、每月工时限量及销售利润见表 2.49。根据市场情况,C 型每月产量应控制在 150

架以内。现工厂需拟定使总利润 z 最大的生产计划。

表 2.49

	A 型	B 型	C 型	工时限量(h)
机身制造时间(h)	1	2	7	2 500
零件装配时间(h)	2	3.5	1	3 500
检验包装时间(h)	1	2	3	1 500
销售利润(元/架)	10	20	50	

(1)写出此问题的数学模型(变量 x_1, x_2, x_3 分别表示 A、B、C 型产品的月产量,约束条件依次为:机身制造时间、零件装配时间、检验包装时间及 C 型产量)。

(2)表 2.50 是用单纯形法解此问题过程中的一个不完全表,请将表完成。

表 2.50

C_B	C X_B	$B^{-1}b$	10 x_1	20 x_2	50 x_3	0 x_4	0 x_5	0 x_6	0 x_7
		400		0	0	1	0	-1	-4
		1 512.5		0	0	0	1	-1.75	4.25
		525		1	0	0	0	0.5	-1.5
		150		0	1	0	0	0	1
检验数									

(3)上表是否为终表?若是,请写出最优生产计划、最大销售利润及检验包装时间的影子价格。

(4)写出此问题的对偶问题数学模型,并指出其最优解。

(5)现又设计出 D 型产品,每架需机身制造、零件装配和检验包装时间分别为 6、2、2(h),销售利润为 30 元/架,问是否应考虑生产 D 型产品?为什么?

2.9 设线性规划问题 1 是

$$\max z_1 = CX$$

$$\begin{cases} AX \leq b \\ X \geq 0 \end{cases}$$

Y^* 是其对偶问题的最优解;又设线性规划问题 2 是

$$\max z_2 = CX$$

$$\begin{cases} AX \leq b + k \\ X \geq 0 \end{cases}$$

式中 k 是已知常向量。求证: $\max z_2 \leq \max z_1 + Y^* k$ 。

2.10 甲、乙、丙三个工厂生产的某种产品产量分别为 7、9、7 个单位,运往 A、B、C、D 四个门市部,各门市部需要量分别为 3、5、7、8 个单位。已知单位运价如表 2.51。试求运输计划使总运费最小。

表 2.51

		门市部			
		A	B	C	D
工厂	甲	12	13	10	11
	乙	10	12	14	10
	丙	14	11	15	12

2.11 已知一运输问题由表 2.52 给出,现有一调运方案由方案表 2.53 给出。

表 2.52

运 价 产 地	销 地				产量(t)
	B ₁	B ₂	B ₃	B ₄	
A ₁	5	10	20	7	3
A ₂	7	2	0	1	6
A ₃	10	3	3	4	7
销 量	6	5	2	3	16

表 2.53

	B ₁	B ₂	B ₃	B ₄
A ₁	3			
A ₂		1	2	3
A ₃	3	4		

(1) 所给方案是否可行? 为什么?

(2) 所给方案是否最优? 若从 x_{21} 空格进行调整, 最大调整量是多少? 可使总运费下降多少?

(3) 若表中给出的单位运价换成单位收入, 要求收入最大的运输方案, 如何利用原表上作业法求解。

(4) 列出此运输问题的线性规划模型。

表 2.54

资 源	型 号		
	小	中	大
尼龙绸	1.5	1.7	1.8
尼龙棉	1.3	1.5	1.6
劳动力	4	4.5	5
设备	2.8	3.8	4.2

2.12 某服装厂制造大、中、小三种尺寸的防寒服, 所用资源有尼龙绸、尼龙棉、劳动力和缝纫设备。缝制一件防寒服所需各种资源的数量如表 2.54 所示(单位已适当给定)。不考虑固定费用, 每种防寒服售出一件所得利润分别为 10、12、13 元, 可用面料尼龙绸 1 500 m, 尼龙棉 1 000 m, 劳动力 3 500 人, 设备 2 800 h。此外, 每种防寒服不管缝制多少件, 只要做都要支付一定的固定费用, 小号为 120 元, 中号为 150 元, 大号为 180 元。现欲制定一生产计划使获得的利润为最大, 写出数学模型。

第3章 非线性规划

在上一章中已经看到,实际中有很多优化问题可以抽象成线性规划模型求解。线性规划模型中,目标函数与约束条件都是决策变量的线性函数。但在实际中也有许多优化问题的数学模型不能完全用线性表达式表示。目标或约束中含有非线性函数的优化问题称为非线性规划问题。本章将讨论非线性规划模型及其求解方法。

3.1 基本概念

3.1.1 非线性规划问题与模型

1. 非线性规划问题

在工程技术和经济管理中广泛存在着非线性规划问题。

例 3.1 (构件表面积问题)要设计一个如图 3.1 所示的半球形和圆柱形相连接的构件,要求在体积一定的条件下确定构件的尺寸,使其表面积最小。

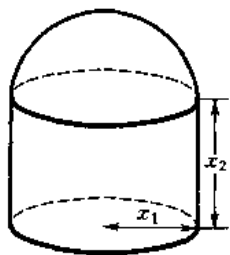


图 3.1

构件的大小取决于其中圆柱体的底半径和高,今设该圆柱的底半径为 x_1 、高为 x_2 。由于构件的表面由半球顶面、侧面和底面组成,因此表面积

$$S = 2\pi x_1^2 + 2\pi x_1 x_2 + \pi x_1^2$$

构件的体积为半球体和圆柱体之和,所以要使构件的体积为定值 V ,应该满足条件

$$\frac{2}{3}\pi x_1^3 + \pi x_1^2 x_2 = V$$

又构件的底半径和圆柱体的高显然非负,故还要求

$$x_1 \geq 0, x_2 \geq 0$$

综合以上分析,该构件表面积最小化问题可归结为如下数学模型:

$$\begin{aligned} \min S &= 3\pi x_1^2 + 2\pi x_1 x_2 \\ \text{s. t. } &\begin{cases} \frac{2}{3}\pi x_1^3 + \pi x_1^2 x_2 = V \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

例 3.2 (生产成本问题)设要在产量不低于某水平 Q_0 的条件下,使生产成本极小化。假设所考虑的决策变量为资本和劳动力,分别记为 x_1 和 x_2 ,则由经济学中的 Cobb-Douglas 生产函数得

$$Q = Ax_1^\alpha x_2^\beta$$

式中, Q 为产出产量; A 为生产技术水平; α 和 β 为参数。又设已知工资率为 ω , 资本报酬率为 γ , 则生产成本

$$C = \gamma x_1 + \omega x_2$$

本问题的数学模型为

$$\begin{aligned} \min & C - \gamma x_1 + \omega x_2 \\ \text{s. t.} & \begin{cases} Ax_1^\alpha x_2^\beta \geq Q_0 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

例 3.3 (投资决策问题) 某公司有 $n (\geq 2)$ 个项目可供选择投资, 并且至少要对其中一个项目投资。已知该公司拥有总资金 A 元, 投资于第 $i (i=1, \dots, n)$ 个项目需花资金 a_i 元, 并预计可收益 b_i 元, 试选择最佳投资方案。

设投资决策变量为

$$x_i = \begin{cases} 1, & \text{决定投资第 } i \text{ 个项目} \\ 0, & \text{决定不投资第 } i \text{ 个项目} \end{cases} \quad (i=1, \dots, n)$$

则投资总金额为 $\sum_{i=1}^n a_i x_i$, 投资总收益为 $\sum_{i=1}^n b_i x_i$, 因为该公司至少要对一个项目投资, 并且总的投资金额不能超过总资金 A , 故有限制条件

$$0 < \sum_{i=1}^n a_i x_i \leq A$$

另外, 由于 $x_i (i=1, \dots, n)$ 只取值 0 或 1, 所以还有

$$x_i(1-x_i)=0 \quad (i=1, \dots, n)$$

最佳投资方案应是投资额最少而总收益最大的方案。因此, 本问题的数学模型为

$$\begin{aligned} \max & R = \frac{\sum_{i=1}^n b_i x_i}{\sum_{i=1}^n a_i x_i} \\ \text{s. t.} & \begin{cases} 0 < \sum_{i=1}^n a_i x_i \leq A \\ x_i(1-x_i)=0 \quad (i=1, \dots, n) \end{cases} \end{aligned}$$

例 3.4 (定位问题) 某连锁超市要选定一个配送中心, 由它向该地区的 n 个超市商场送货。假设该地的所有道路均是相互平行或垂直的 (图 3.2), 试确定这个配送中心的位置, 使它到每个商场的距离不超过 d , 且它到各商场的总距离为最短。

在地面上引进直角坐标系如图 3.2, 设配送中心的位置为 $C(x_1, x_2)$, n 个商场的坐标为 $P_i(a_i, b_i) (i=1, \dots, n)$, 则配送中心到第 i 个用户的距离为

$$|x_1 - a_i| + |x_2 - b_i| \quad (i=1, \dots, n)$$

此问题的数学模型为

$$\begin{aligned} \min & D = \sum_{i=1}^n [|x_1 - a_i| + |x_2 - b_i|] \\ \text{s. t.} & |x_1 - a_i| + |x_2 - b_i| \leq d \quad (i=1, \dots, n) \end{aligned}$$

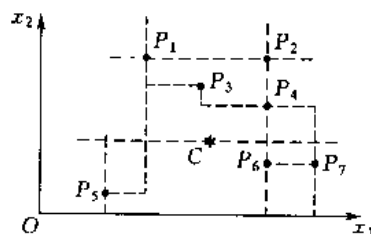


图 3.2

2. 非线性规划模型

由上面的例子可以看出, 非线性规划问题的数学模型也是由决策变量、目标函数和约束条件三要素组成的, 只不过模型的目标或约束中一般含有非线性表达式。为了写出模型的一般

形式,记决策变量 $\mathbf{X} = [x_1 \cdots x_n]^T$, 目标函数 $f(\mathbf{X}) = f(x_1, \cdots, x_n)$, 不等式的约束条件 $g_j(\mathbf{X}) = g_j(x_1, \cdots, x_n) \geq 0 (j = 1, \cdots, l)$, 等式约束条件 $h_i(\mathbf{X}) = h_i(x_1, \cdots, x_n) = 0 (i = 1, \cdots, m)$, 则可写出非线性规划模型的一般式为

$$\begin{aligned} & \min f(\mathbf{X}) \\ (\text{NLP}) \text{ s. t. } & \begin{cases} h_i(\mathbf{X}) = 0 & (i = 1, \cdots, m) \\ g_j(\mathbf{X}) \geq 0 & (j = 1, \cdots, l) \end{cases} \end{aligned}$$

对于目标函数为 $\max f(\mathbf{X})$ 的情形, 可通过转化为 $\min(-f(\mathbf{X}))$ 的形式化为极小型; 对于约束条件为“ \leq ”不等式的情形, 可通过两端乘以“-1”化为“ \geq ”的形式。若记

$$D = \{\mathbf{X} \in \mathbf{R}^n \mid h_i(\mathbf{X}) = 0, i = 1, \cdots, m; g_j(\mathbf{X}) \geq 0, j = 1, \cdots, l\}$$

则(NLP)也可表示为

$$\min_{\mathbf{X} \in D} f(\mathbf{X})$$

其中 D 称为(NLP)的约束集或可行域。当 $D = \mathbf{R}^n$ 时, (NLP)称做无约束极值问题; 当 $D \neq \mathbf{R}^n$ 时, (NLP)称为约束极值问题。

3.1.2 模型的解及相关概念

1. 可行解与最优解

约束集 D 中的点 \mathbf{X} 称为(NLP)的可行解。如果有 $\mathbf{X}^* \in D$, 对于任意的 $\mathbf{X} \in D$, 都有 $f(\mathbf{X}^*) \leq f(\mathbf{X})$, 则称 \mathbf{X}^* 为(NLP)的最优解, 也称全局极小值点; 如果对于 $\mathbf{X}^0 \in D$, 使得在 \mathbf{X}^0 的邻域

$$B(\mathbf{X}^0, \epsilon) = \{\mathbf{X} \mid \|\mathbf{X} - \mathbf{X}^0\| < \epsilon\}$$

中的任意 $\mathbf{X} \in D$, 都有 $f(\mathbf{X}^0) \leq f(\mathbf{X})$, 则称 \mathbf{X}^0 为(NLP)的一个局部极小值点, 其中 $\epsilon > 0$, $\|\cdot\|$ 表示 n 维向量的模, 通常定义为 n 个分量平方和的算术根。

在上述定义中, 如果将不等号反向, 则可得到极大值点的定义。极小值点和极大值点可简称为极小点和极大点。

下面举一个简单例子给出可行解与最优解的图示。

例 3.5 考虑非线性规划问题

$$\min f(\mathbf{X}) = (x_1 - 2)^2 + (x_2 - 2)^2$$

$$\text{s. t. } x_1 + x_2 = 6$$

目标与约束的图形如图 3.3 所示。

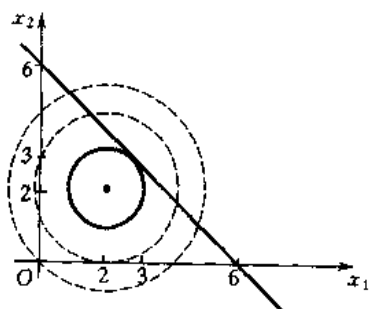


图 3.3

由图易见, 可行解是位于直线 $x_1 + x_2 = 6$ 上的点, 目标是一族以 $(2, 2)$ 为圆心的圆, 最优解是圆族中与 $x_1 + x_2 = 6$ 相切的圆的切点 $(3, 3)$ 。

在本例中若将约束改为 $x_1 + x_2 \leq 6$, 则可行解是位于直线 $x_1 + x_2 = 6$ 左下半平面上的点, 最优解为圆心 $(2, 2)$ 。这时的约束实际上不起作用。

已知线性规划的最优解一定在可行域的边界(特别是顶点)上达到, 而本例说明, 非线性规划的最优解可以在可行域的任意点达到。

2. 梯度、海赛阵与泰勒公式

若 $f(\mathbf{X})$ 在 \mathbf{X}_0 的邻域内有连续一阶偏导数, 则称 $f(\mathbf{X})$ 在 \mathbf{X}_0 点对 n 个变元的偏导数组成的向量为 $f(\mathbf{X})$ 在 \mathbf{X}_0 的梯度, 记为 $\nabla f(\mathbf{X}_0)$, 即

$$\nabla f(\mathbf{X}_0) = \left[\frac{\partial f(\mathbf{X}_0)}{\partial x_1} \quad \dots \quad \frac{\partial f(\mathbf{X}_0)}{\partial x_n} \right]^T$$

梯度的几何意义是过 \mathbf{X}_0 点且与 $f(\mathbf{X})$ 在 \mathbf{X}_0 的切平面垂直的向量, 梯度向量的方向是函数值在该点增加最快的方向。

若 $f(\mathbf{X})$ 在 \mathbf{X}_0 的邻域内有连续二阶偏导数, 则称 $f(\mathbf{X})$ 在 \mathbf{X}_0 点对 n 个变元两两组合的二阶偏导数组成的矩阵为 $f(\mathbf{X})$ 在 \mathbf{X}_0 的海赛(Hesse)阵, 记为 $\mathbf{H}_f(\mathbf{X}_0)$, 或简记为 $\mathbf{H}(\mathbf{X}_0)$, 即

$$\mathbf{H}(\mathbf{X}_0) = \left[\frac{\partial^2 f(\mathbf{X}_0)}{\partial x_i \partial x_j} \right]_{n \times n} = \begin{bmatrix} \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_1^2} & \dots & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f(\mathbf{X}_0)}{\partial x_n^2} \end{bmatrix}$$

若 $f(\mathbf{X})$ 在 \mathbf{X}_0 的邻域内有连续二阶偏导数, 则可写出 $f(\mathbf{X})$ 在 \mathbf{X}_0 的(二阶)泰勒(Taylor)展开式

$$f(\mathbf{X}) = f(\mathbf{X}_0) + \nabla f(\mathbf{X}_0)^T (\mathbf{X} - \mathbf{X}_0) + \frac{1}{2} (\mathbf{X} - \mathbf{X}_0)^T \mathbf{H}(\mathbf{X}_0) (\mathbf{X} - \mathbf{X}_0) + o(\|\mathbf{X} - \mathbf{X}_0\|^2)$$

其中 $o(\|\mathbf{X} - \mathbf{X}_0\|^2)$ 是当 $\mathbf{X} \rightarrow \mathbf{X}_0$ 时 $\|\mathbf{X} - \mathbf{X}_0\|^2$ 的高阶无穷小。

例 3.6 写出 $f(\mathbf{X}) = 3x_1^2 + \sin x_2$ 在 $\mathbf{X}_0 = [0 \ 0]^T$ 点的二阶泰勒展开式。

解 $f(\mathbf{X})$ 显然有连续的二阶偏导数, 其梯度

$$\nabla f(\mathbf{X}) = [6x_1 \quad \cos x_2]^T, \quad \nabla f(\mathbf{X}_0) = [0 \ 1]^T$$

其海赛阵

$$\mathbf{H}(\mathbf{X}) = \begin{bmatrix} 6 & 0 \\ 0 & -\sin x_2 \end{bmatrix}, \quad \mathbf{H}(\mathbf{X}_0) = \begin{bmatrix} 6 & 0 \\ 0 & 0 \end{bmatrix}$$

在 \mathbf{X}_0 点的二阶泰勒展开式为

$$\begin{aligned} f(\mathbf{X}) &= 0 + [0 \ 1] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} [x_1 \ x_2] \begin{bmatrix} 6 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + o(\|\mathbf{X}\|^2) \\ &= x_2 + 3x_1^2 + o(\|\mathbf{X}\|^2) \end{aligned}$$

如果忽略 $o(\|\mathbf{X}\|^2)$, 则 $f(\mathbf{X})$ 在 \mathbf{X}_0 点的近似表达式为

$$f(\mathbf{X}) = 3x_1^2 + x_2$$

3. 极值的条件

对于无约束极值问题, 可以利用微积分中的知识给出局部极值点的条件。将 n ($n > 1$) 元函数 $f(\mathbf{X})$ 与一元函数 $f(x)$ 的极值条件加以对比并归纳于表 3.1。

表 3.1

\mathbf{X}_0 (或 x_0) 是极小值点	必要条件	充分条件
一元函数 $f(x)$	$f'(x_0) = 0$	$f'(x_0) = 0$, 且 $f''(x_0) > 0$
n 元函数 $f(\mathbf{X})$	$\nabla f(\mathbf{X}_0) = \mathbf{0}$	$\nabla f(\mathbf{X}_0) = \mathbf{0}$ 且 $\mathbf{H}(\mathbf{X}_0) > 0$

其中的充分条件主要依据函数的二阶导信息, $H(X_0) > 0$ 表示海赛阵正定。根据线性代数的知识可知, 如果一个方阵的各阶主子式均大于零, 则可判定该方阵为正定的。若 X_0 是极大值点, 则充分条件中不等号反向。

利用极值的条件, 可以求一些无约束极值问题的极值点。即先由必要条件求出使 $\nabla f(X_0) = 0$ 的点 X_0 (称为驻点), 再由充分条件判断 X_0 的极值性。

例 3.7 求 $f(X) = 2x_1^2 - 8x_1 + 2x_2^2 - 4x_2 + 20$ 的极小值点。

解 $\nabla f(X) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}^T = [4x_1 - 8 \quad 4x_2 - 4]^T$

令 $\nabla f(X) = 0$, 解得驻点 $X_0 = [2 \quad 1]^T$, 又 $H(X) = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$, 显然 $H(X_0) > 0$, 故 X_0 是极小值点。

4. 凸规划

在非线性规划模型(NLP)中, 若目标函数 $f(X)$ 是凸函数, 不等式约束函数 $g_j(X)$ 为凹函数 ($j = 1, \dots, l$), 等式约束函数 $h_i(X)$ 为仿射函数 ($i = 1, \dots, m$), 则称(NLP)为一个凸规划。

在微积分中已知, 凸函数 $f(X)$ 是定义在凸集 D 上且满足对任意 $X_1, X_2 \in D, \alpha \in [0, 1]$ 有下式成立的函数:

$$f(\alpha X_1 + (1 - \alpha)X_2) \leq \alpha f(X_1) + (1 - \alpha)f(X_2)$$

若不等式中严格不等号成立, 则 $f(X)$ 为严格凸函数。凸函数的几何形状是下单峰, 因此它的局部极值点就是全局极值点。如果 $f(X)$ 是凸函数, 则 $-f(X)$ 就是凹函数, 凹函数的几何形状是上单峰。仿射函数即线性函数, 它既可看做是凸函数, 又可看做是凹函数。仿射函数的几何形状是平面或直线。

判定一个函数 $f(X)$ 是否凸函数, 除根据定义外, 还可利用 $f(X)$ 的导数信息。在微积分中已知, 若一元函数 $f(x)$ 的二阶导大于等于零, 则 $f(x)$ 是凸函数。对于多元函数 $f(X)$, 若其海赛阵 $H(X)$ 半正定, 则 $f(X)$ 是凸函数。对于凹函数 $f(X)$ 的判定, 可化为 $-f(X)$ 的凸函数判定, 也可直接判定。若 $f(X)$ 的海赛阵 $H(X)$ 半负定, 即 $H(X)$ 的奇数阶主子式均为非正而偶数阶主子式均为非负, 则 $f(X)$ 为凹函数。

例 3.8 判断下面的非线性规划是否为凸规划;

$$\begin{aligned} \max \quad & f(X) = x_1 + x_2 \\ \text{s. t.} \quad & \begin{cases} x_1^2 + x_2^2 \leq 1 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned}$$

解 模型可写为

$$\begin{aligned} \min \quad & (-f(X)) = -x_1 - x_2 \\ \text{s. t.} \quad & \begin{cases} g_1(X) = 1 - x_1^2 - x_2^2 \geq 0 \\ g_2(X) = x_1 \geq 0 \\ g_3(X) = x_2 \geq 0 \end{cases} \end{aligned}$$

$$H_{-f}(X) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \geq 0, H_{g_1}(X) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} > 0,$$

$$H_{-g_2}(X) = H_{-g_3}(X) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \geq 0$$

说明 $-f(X)$ 是凸函数, $g_1(X)$ 、 $g_2(X)$ 、 $g_3(X)$ 是凹函数。因此, 本模型是一个凸规划。

由定义不难推出凸规划的一些简单性质: 凸规划的约束集是凸集; 凸规划的最优解集也是凸集; 任何局部最优解也是全局最优解; 若目标函数为严格凸函数, 且最优解存在, 则其最优解是惟一的。

容易想到, 线性规划也是一种凸规划。

3.2 无约束极值问题

无约束极值问题是非线性规划的一种类型。它不仅在实际中有应用, 而且是研究约束极值问题的基础之一。因此, 研究无约束极值问题的求解方法具有重要意义。

无约束极值问题的一般表达式为

$$\min f(X)$$

其中 $X \in \mathbf{R}^n$ 。当 $f(X)$ 为可微函数时, 可以利用极值的条件求解。但由条件 $\nabla f(X) = 0$ 得到的常常是一个非线性方程组, 求解它十分困难。因此, 求解无约束极值问题一般是采用迭代法, 称为下降类算法。

3.2.1 下降类算法的基本步骤与算法的收敛性

求解无约束极值问题的下降类算法的基本思想是使 $f(X)$ 逐步下降, 逐渐趋近其极小值。迭代的方式是从某一初始点 X_0 出发, 选取某一搜索方向 P_0 , 沿该方向搜索到下一个点 X_1 。若其已达到与最优解误差的精度要求, 则停止, 否则再沿该点的某一方向 P_1 搜索下一个点 X_2 。这一过程的图示如图 3.4。方法的一般步骤如下:

①选取初始点 X_0 , 令 $k := 0$ ($:=$ 为赋值号, $k := 0$ 表示把 0 赋给变量 k), 确定精度 $\varepsilon > 0$;

②对于点 X_k , 计算 $\nabla f(X_k)$, 若 $\|\nabla f(X_k)\| < \varepsilon$, 则停止, 得到近似最优解 X_k , 否则转③;

③从 X_k 出发, 确定搜索方向 P_k ;

④沿 P_k 方向搜索, 即由 $X = X_k + \lambda P_k$ 确定搜索步长 λ_k , 得下一点 $X_{k+1} = X_k + \lambda_k P_k$, 令 $k := k + 1$, 转②。

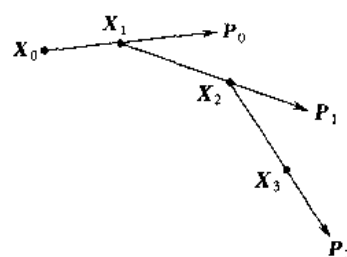


图 3.4

在上述步骤中, 方向 P_k 的选定具有关键作用, 不同形式的 P_k 便形成了不同的算法, 而不同的算法所产生的点列 $\{X_k\}$ 收敛于最优解 X^* 的速度也不相同。衡量算法的收敛速度, 一般是看迭代点列中相邻两项与 X^* 的距离之比的极限

$$\lim_{k \rightarrow \infty} \frac{\|X_{k+1} - X^*\|}{\|X_k - X^*\|^\alpha} = \beta$$

当 $\alpha = 1$ 时, 显然有 $\beta \in [0, 1]$ 。如果 $0 < \beta < 1$, 则称 $\{X_k\}$ 是线性收敛的; 如果 $\beta = 0$, 则称 $\{X_k\}$ 是超线性收敛的。当 $\alpha = 2$ 时, 如果 $0 \leq \beta < +\infty$, 则称 $\{X_k\}$ 是二阶收敛的。在这三种收敛速度中, 以二阶收敛速度最快, 超线性收敛速度次之。

搜索方向 P_k 确定后,沿 P_k 搜寻下一个迭代点的工作主要是由 $X = X_k + \lambda P_k$ 确定步长 λ_k 。由于这里只涉及一个变元 λ 的线性表达式,故称一维搜索或线性搜索。

上述步骤②中的 $\|\nabla f(X_k)\| < \epsilon$ 是一种常用的收敛准则,是基于极值的条件 $\nabla f(X^*) = 0$ 得到的。其他收敛准则还有 $\|X_{k+1} - X_k\| < \epsilon$ 等。

3.2.2 一维搜索

一维搜索即由 $X = X_k + \lambda P_k$ 确定步长 λ_k ,其方法可分为精确方法和近似方法两类。精确方法是指通过求极值

$$\min_{\lambda \geq 0} f(X_k + \lambda P_k)$$

的精确值来得到相应极值点 λ_k 。这样得到的 λ_k 称做最佳步长。近似方法是通过求上述极值的近似值来得到 λ_k 的方法。常用的有不用导数信息的分数法和 0.618 法,以及用到导数信息的近似最佳步长公式等。

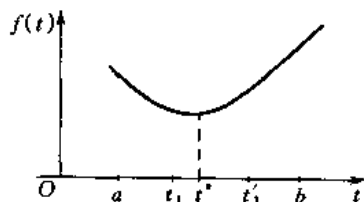


图 3.5

1. 分数法与 0.618 法

分数法又称斐波那契(Fibonacci)法,是一种求一元函数 $f(t)$ 近似极小点的试探方法。其基本思想是,由于 $f(t)$ 在包含极小点 t^* 的足够小的区间 $[a, b]$ 上可以认为是下单峰的,故在 $[a, b]$ 内任取二点 $t_1 < t'_1$,若 $f(t_1) < f(t'_1)$,则必有 $t^* \in [a, t'_1]$ (图 3.5),可去掉区间 $[t'_1, b]$ 。再在新区间 $[a, t'_1]$ 中取一点 $t_2 > t_1$,若 $f(t_2) > f(t_1)$,说明 $t^* \in [a, t_2]$,可去掉 $[t_2, t'_1]$ 。这样反复取点、比较,缩小含 t^* 的区间,直至最后所得的区间足够小,则其中任何一点可作为 t^* 的近似值。

衡量区间“足够小”的方法有两种,即采用绝对精度或相对精度。记第 n 次缩短后的小区间为 $[a_n, b_n]$,给定的精度要求为 $\epsilon > 0$,则满足绝对精度是指

$$|b_n - a_n| < \epsilon$$

满足相对精度是指

$$\frac{|b_n - a_n|}{|b - a|} < \epsilon$$

那么,怎样在区间中取点呢?可以证明,按斐波那契数构造取点比例对称取点效果最好。这就是分数法。

记斐波那契数的一般项为 F_n 。其表达式为

$$\begin{cases} F_n = F_{n-1} + F_{n-2} \\ F_0 = F_1 = 1 \end{cases}$$

表 3.2 给出 F_n 前 13 项的值。

表 3.2

n	0	1	2	3	4	5	6	7	8	9	10	11	12
F_n	1	1	2	3	5	8	13	21	34	55	89	144	233

如果按 $\frac{F_{n-1}}{F_n}$ 的比例在区间中对称地取点 t_1 和 t'_1 (图 3.6),经比较后去掉 $[t'_1, b]$,则 t_1

在 $[a, t'_1]$ 中所占的比例为

$$\frac{(1 - \frac{F_{n-1}}{F_n})(b-a)}{\frac{F_{n-1}}{F_n}(b-a)} = \frac{F_n - F_{n-1}}{F_{n-1}} = \frac{F_{n-2}}{F_{n-1}}$$

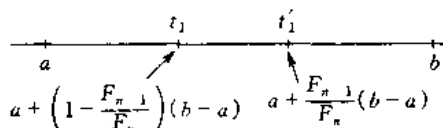


图 3.6

再按 $\frac{F_{n-2}}{F_{n-1}}$ 的比例在 $[a, t'_1]$ 中取 t_2 与 t_1 对称, 这样

反复 $n-1$ 次后所得小区间 $[a_{n-1}, b_{n-1}]$ 长缩短为

$$(b-a) \frac{F_{n-1}}{F_n} \cdot \frac{F_{n-2}}{F_{n-1}} \cdots \frac{F_2}{F_3} \cdot \frac{F_1}{F_2} = \frac{1}{F_n} (b-a)$$

与原区间长之比为

$$\frac{\frac{1}{F_n}(b-a)}{b-a} = \frac{1}{F_n}$$

恰为相对精度表达式。于是, 在开始时可由相对精度 $\frac{1}{F_n} < \epsilon$ 解出 F_n 和 n 。可将上述的分数法归纳为下列步骤:

- ①由所给精度 ϵ 和 $\frac{1}{F_n} < \epsilon$ 定 n ;
- ②由 $\frac{F_{n-1}}{F_n}$ 的比例在区间 $[a, b]$ 内对称取点 $t_1 = a + (1 - \frac{F_{n-1}}{F_n})(b-a)$, $t'_1 = a + \frac{F_{n-1}}{F_n}(b-a)$, 比较 $f(t_1)$ 和 $f(t'_1)$, 去掉 $\max\{f(t_2), f(t'_1)\}$ 相应点以外的区间, 得新区间 $[a_1, b_1]$;
- ③在 $[a_1, b_1]$ 中按 $\frac{F_{n-2}}{F_{n-1}}$ 的比例取点 t_2 (与区间中所剩的点对称), 比较 $f(t_2)$ 和 $\min\{f(t_1), f(t'_1)\}$, 去掉较大者相应点以外的区间, 得新区间 $[a_2, b_2]$;
- ④在 $[a_2, b_2]$ 中按 $\frac{F_{n-3}}{F_{n-2}}$ 的比例取点 t_3 , 再比较、缩短, 直至取完 $\frac{F_1}{F_2}$, 最后的小区间 $[a_{n-1}, b_{n-1}]$ 中任一点可作为 t^* 的近似值。

例 3.9 用分数法求 $f(t) = t^2 - t + 2$ 在区间 $[-1, 3]$ 上的近似极小点, 要求缩短后的区间长度不大于原区间长的 8%。

解 由于 $f''(t) = 2 > 0$, 故 $f(t)$ 是严格凸函数, 由 $f'(t) = 2t - 1 = 0$ 解得的 $t^* = \frac{1}{2}$ 是极小点, $f(t^*) = 1.75$ 。

由 $\frac{1}{F_n} \leq 0.08$ 知 $F_n \geq \frac{1}{0.08} = 12.5$, 查表 3.2 得 $n = 6$, $\frac{F_5}{F_6} = \frac{8}{13}$

$$t_1 = -1 + (1 - \frac{8}{13})[3 - (-1)] = 0.538, f(t_1) = 1.751$$

$$t'_1 = -1 + \frac{8}{13}[3 - (-1)] = 1.462, f(t'_1) = 2.675 > f(t_1)$$

故 $[a_1, b_1] = [-1, 1.462]$, $\frac{F_4}{F_5} = \frac{5}{8}$

$$t_2 = -1 + (1 - \frac{5}{8})[1.462 - (-1)] = -0.077$$

$$f(t_2) = 2.083 > f(t_1), \text{故 } [a_2, b_2] = [-0.077, 1.462]$$

依此进行,最后得 $t_5 = 0.538$ 为近似极小点, $f(t_5) = 1.751$ 为近似极小值。

因分数法每次取点比例都在变而计算较繁,所以提出分数法的近似方法——0.618 法。该方法与分数法的区别仅在于,每次取点的比例都是定值 0.618,即每次区间内二点的位置均在区间相对长度的 0.382 和 0.618 处。由于 0.618 法的效果也较好,且更加简便,因而更易于应用。

2. 近似最佳步长公式

近似最佳步长公式与用精确方法求最佳步长的基本思想一致,都是通过求 $f(\mathbf{X}_k + \lambda \mathbf{P}_k)$ 关于 λ 的极小点来得到步长 λ_k ,只是近似最佳步长公式先将 $f(\mathbf{X}_k + \lambda \mathbf{P}_k)$ 用其泰勒展开式近似表示,以简化求极值点的计算。

设 $f(\mathbf{X})$ 存在连续的二阶偏导数,则有泰勒展开式

$$f(\mathbf{X}_k + \lambda \mathbf{P}_k) \approx f(\mathbf{X}_k) + \nabla f(\mathbf{X}_k)^T \lambda \mathbf{P}_k + \frac{1}{2} \lambda^2 \mathbf{P}_k^T \mathbf{H}(\mathbf{X}_k) \mathbf{P}_k$$

对 λ 求导,并令导数为 0,得

$$\frac{df}{d\lambda} \approx \nabla f(\mathbf{X}_k)^T \mathbf{P}_k + \lambda \mathbf{P}_k^T \mathbf{H}(\mathbf{X}_k) \mathbf{P}_k = 0$$

解得

$$\lambda = \frac{-\nabla f(\mathbf{X}_k)^T \mathbf{P}_k}{\mathbf{P}_k^T \mathbf{H}(\mathbf{X}_k) \mathbf{P}_k}$$

此式即为 λ_k 的近似最佳步长公式。当 $f(\mathbf{X})$ 为二次函数时,此公式是精确的。

例 3.10 设 $f(\mathbf{X}) = (x_1 - 1)^2 + (x_2 - 1)^2$, $\mathbf{X}_k = [0 \ 0]^T$, $\mathbf{P}_k = -\nabla f(\mathbf{X}_k)$, 用近似最佳步长公式求 λ_k 。

解 $\nabla f(\mathbf{X}) = [2(x_1 - 1) \ 2(x_2 - 1)]^T$, $-\nabla f(\mathbf{X}_k) = [-2 \ -2]^T$

$$\mathbf{H}(\mathbf{X}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \mathbf{H}(\mathbf{X}_k)$$

$$\lambda_k = -\frac{[-2 \ -2] \begin{bmatrix} -2 \\ -2 \end{bmatrix}}{[-2 \ -2] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix}} = \frac{1}{2}$$

由于这里的 $f(\mathbf{X})$ 是二次函数,故所求 λ_k 是精确的。

3.2.3 梯度法与共轭梯度法

1. 梯度法

梯度法是一种求解无约束极值问题的最简单、最基本的下降类算法。该方法选取搜索方向 \mathbf{P}_k 的出发点是:怎样选取 \mathbf{P}_k 可使 $f(\mathbf{X})$ 下降得最快?或者说使

$$f(\mathbf{X}_k + \lambda \mathbf{P}_k) - f(\mathbf{X}_k) < 0$$

且不等式左式的绝对值尽量大。

设 $f(\mathbf{X})$ 有连续的二阶偏导数, 则有泰勒展开式

$$f(\mathbf{X}_k + \lambda \mathbf{P}_k) \approx f(\mathbf{X}_k) + \lambda \nabla f(\mathbf{X}_k)^\top \mathbf{P}_k$$

或

$$f(\mathbf{X}_k + \lambda \mathbf{P}_k) - f(\mathbf{X}_k) \approx \lambda \nabla f(\mathbf{X}_k)^\top \mathbf{P}_k$$

由于 $\lambda > 0$, 应使

$$\nabla f(\mathbf{X}_k)^\top \mathbf{P}_k = \|\nabla f(\mathbf{X}_k)\| \cdot \|\mathbf{P}_k\| \cos \theta < 0$$

其中 θ 为向量 $\nabla f(\mathbf{X}_k)$ 和 \mathbf{P}_k 的夹角。为使上式成立, 应有 $\theta > \frac{\pi}{2}$ 。又为使绝对值尽量大, 应使 $\cos \theta = -1$, 即 $\theta = \pi$ 。这说明

$$\mathbf{P}_k = -\nabla f(\mathbf{X}_k)$$

即在 \mathbf{X}_k 点 $f(\mathbf{X})$ 的负梯度方向是使 $f(\mathbf{X})$ 下降最快的方向。以负梯度方向作为搜索方向 \mathbf{P}_k 的下降类算法称为梯度法或最速下降法。其一般步骤如下:

①选取初始点 \mathbf{X}_0 , 令 $k := 0$, 确定精度 $\epsilon > 0$;

②对于点 \mathbf{X}_k , 计算 $\nabla f(\mathbf{X}_k)$, 若 $\|\nabla f(\mathbf{X}_k)\| < \epsilon$, 则停止计算, 得到近似最优解 \mathbf{X}_k , 否则转③;

③取 $\mathbf{P}_k = -\nabla f(\mathbf{X}_k)$, 沿 \mathbf{P}_k 进行一维搜索, 得到 λ_k 和下一点 $\mathbf{X}_{k+1} = \mathbf{X}_k + \lambda_k \mathbf{P}_k$;

④计算 $\nabla f(\mathbf{X}_{k+1})$, 置 $k := k + 1$, 转②。

例 3.11 用梯度法求 $f(\mathbf{X}) = (x_1 - 1)^2 + (x_2 - 1)^2$ 的极小点, 取 $\epsilon = 0.1$, 初始点 $\mathbf{X}_0 = [0 \ 0]^\top$ 。

解 $\nabla f(\mathbf{X}) = [2(x_1 - 1) \ 2(x_2 - 1)]^\top$, $\nabla f(\mathbf{X}_0) = (-2, -2)^\top$

$$\|\nabla f(\mathbf{X}_0)\| = \sqrt{(-2)^2 + (-2)^2} = \sqrt{8} > \epsilon$$

由近似最佳步长公式求 λ_0 :

$$\mathbf{H}(\mathbf{X}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \lambda_0 = \frac{[-2 \ -2] \begin{bmatrix} -2 \\ -2 \end{bmatrix}}{[-2 \ -2] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix}} = \frac{1}{2}$$

$$\mathbf{X}_1 = \mathbf{X}_0 - \lambda_0 \nabla f(\mathbf{X}_0) = [0 \ 0]^\top - \frac{1}{2} [-2 \ -2]^\top = [1 \ 1]^\top$$

$$\nabla f(\mathbf{X}_1) = [0 \ 0]^\top, \|\nabla f(\mathbf{X}_1)\| = 0 < \epsilon$$

故

$$\mathbf{X}_1 = [1 \ 1]^\top = \mathbf{X}^*$$

在本例中, 目标函数 $f(\mathbf{X})$ 是同心圆族。无论初始点选在何处, $f(\mathbf{X})$ 在该点的负梯度方向总是指向圆心, 而圆心就是极小点, 故沿负梯度方向搜索一步便可得极小点。

但对于一般的函数, 若每次迭代均采用负梯度方向, 则由于这些方向是彼此正交的, 很可能形成开头几步函数值下降较快, 但后来便产生直角锯齿状的“拉锯”现象, 收敛速度很慢。事实上可以证明, 梯度法是线性收敛的。为了加快收敛速度, 在梯度法的基础上又发展出了一些改进的方法。

2. 共轭梯度法

首先说明共轭和共轭方向的概念。

设 X 与 Y 均为 n 维向量, A 是 n 阶对称正定阵。若 $X^T A Y = 0$, 则称 X 与 Y 关于 A 共轭。当 $A = I$, 则 X 与 Y 正交。关于共轭, 有下述重要性质。

①若 n 个非零 n 维向量 P_1, \dots, P_n 关于 A 共轭, 则 P_1, \dots, P_n 线性无关。事实上, 只需考虑 $\alpha_1 P_1 + \dots + \alpha_n P_n = 0, \alpha_i \in \mathbb{R}^1$, 两边分别左乘 $P_i^T A (i = 1, \dots, n)$, 则有 $\alpha_i P_i^T A P_i = 0$, 而 A 正定, 即 $P_i^T A P_i > 0$, 故 $\alpha_i = 0$, 说明 P_1, \dots, P_n 线性无关。这一性质也说明不可能有 $n+1$ 个 n 维向量关于 A 共轭。

②若非零 n 维向量 P_1, \dots, P_n 关于 H 共轭, 则以 H 为二次项系数阵的二次函数 $f(X) = \frac{1}{2} X^T H X + C^T X$ 沿 P_1, \dots, P_n (称为共轭方向) 进行精确一维搜索, 至多经 n 次便可得极小点。

这一性质说明采用共轭方向作为搜索方向, 对二次函数求极小可以有限步终止。这是一个很好的性质 (称为二次终止性), 由此可构造二次函数的共轭方向算法。共轭方向算法用于二次函数时均具有二次终止性。由于一般函数在一点附近的性质往往与二次函数很相似, 因此共轭方向算法一般也可用于其他非线性函数, 并且至少是线性收敛的。下面给出一种较著名的共轭方向法——共轭梯度法。其一般步骤如下:

- ①选取初始点 X_0 , 令 $k := 0$, 确定精度 $\epsilon > 0$;
- ②取 $P_0 = -\nabla f(X_0)$;
- ③对 X_k , 求解 $\min_{\lambda} f(X_k + \lambda P_k)$, 得 λ_k 和 $X_{k+1} = X_k + \lambda_k P_k$;
- ④若 $\|\nabla f(X_{k+1})\| < \epsilon$, 停止, $X_{k+1} = X^*$, 否则, 若 $k < n-1$, 转⑤, 若 $k = n-1$, 重置 $X_0 = X_n$, 转②; (这里重置的原因是由于计算中的舍入误差可能造成 n 步未终止。)
- ⑤取 $P_{k+1} = -\nabla f(X_{k+1}) + \beta_k P_k$, 令 $k := k+1$, 转③, 其中

$$\beta_k = \frac{\nabla f(X_{k+1})^T \nabla f(X_{k+1})}{\nabla f(X_k)^T \nabla f(X_k)}$$

可以证明, 按此方向构造的 P_k 是关于 H 共轭的, 并且有 $P_k^T \nabla f(X_{k+1}) = 0$ 。

当 $f(X) = \frac{1}{2} X^T H X + C^T X, \nabla f(X) = HX + C$ 时, 有

$$\nabla f(X_{k+1}) - \nabla f(X_k) = H(X_{k+1} - X_k) = H(X_k + \lambda_k P_k - X_k) = H\lambda_k P_k$$

两边同时左乘 P_k^T , 有

$$-P_k^T \nabla f(X_k) = \lambda_k P_k^T H P_k$$

从而有

$$\lambda_k = \frac{-P_k^T \nabla f(X_k)}{P_k^T H P_k}$$

例 3.12 求 $f(X) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1$ 的极小点, 取 $X_0 = [-2 \ 4]^T$ 。

$$\text{解 } f(X) = \frac{1}{2} [x_1 \ x_2] \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [-2 \ 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\nabla f(X) = [3x_1 - x_2 - 2 \quad x_2 - x_1]^T$$

$$\nabla f(X_0) = [-12 \ 6]^T, P_0 = -\nabla f(X_0) = [12 \ -6]^T$$

$$\lambda_0 = \frac{-\mathbf{P}_0^T \nabla f(\mathbf{X}_0)}{\mathbf{P}_0^T \mathbf{H} \mathbf{P}_0} = \frac{-[12 \quad -6] \begin{bmatrix} -12 \\ 6 \end{bmatrix}}{[12 \quad -6] \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 12 \\ -6 \end{bmatrix}} = \frac{5}{17}$$

$$\mathbf{X}_1 = \mathbf{X}_0 + \lambda \mathbf{P}_0 = [-2 \quad 4]^T + \frac{5}{17} [12 \quad -6]^T = \left[\frac{26}{17} \quad \frac{38}{17} \right]^T$$

$$\nabla f(\mathbf{X}_1) = \left[\frac{6}{17} \quad \frac{12}{17} \right]^T$$

$$\beta_0 = \frac{\nabla f(\mathbf{X}_1)^T \nabla f(\mathbf{X}_1)}{\nabla f(\mathbf{X}_0)^T \nabla f(\mathbf{X}_0)} = \frac{\begin{bmatrix} \frac{6}{17} & \frac{12}{17} \end{bmatrix} \begin{bmatrix} \frac{6}{17} \\ \frac{12}{17} \end{bmatrix}}{[-12 \quad 6] \begin{bmatrix} -12 \\ 6 \end{bmatrix}} = \frac{1}{289}$$

$$\mathbf{P}_1 = -\nabla f(\mathbf{X}_1) + \beta_0 \mathbf{P}_0 = -\left[\frac{6}{17} \quad \frac{12}{17} \right] + \frac{1}{289} [12 \quad -6] = \left[-\frac{90}{289} \quad -\frac{210}{289} \right]^T$$

$$\lambda_1 = \frac{-\mathbf{P}_1^T \nabla f(\mathbf{X}_1)}{\mathbf{P}_1^T \mathbf{H} \mathbf{P}_1} = \frac{17}{10}, \mathbf{X}_2 = \mathbf{X}_1 + \lambda_1 \mathbf{P}_1 = [1 \quad 1]^T$$

由于 $\nabla f(\mathbf{X}_2) = [0 \quad 0]^T$, 故 $\mathbf{X}_2 = [1 \quad 1]^T$ 即极小点, 计算经两步终止。

3.2.4 牛顿法与拟牛顿法

1. 牛顿法

设 $f(\mathbf{X})$ 具有连续的二阶偏导数, 由泰勒公式, $f(\mathbf{X})$ 可以近似地表示为

$$f(\mathbf{X}) = f(\mathbf{X}_k) + \nabla f(\mathbf{X}_k)^T (\mathbf{X} - \mathbf{X}_k) + \frac{1}{2} (\mathbf{X} - \mathbf{X}_k)^T \mathbf{H}(\mathbf{X}_k) (\mathbf{X} - \mathbf{X}_k)$$

从而

$$\nabla f(\mathbf{X}) = \nabla f(\mathbf{X}_k)^T + \mathbf{H}(\mathbf{X}_k) (\mathbf{X} - \mathbf{X}_k)$$

当 $f(\mathbf{X})$ 为二次函数时, 上式为精确表达式。

取 \mathbf{X}_{k+1} 使 $f(\mathbf{X})$ 达极小, 应有 $\nabla f(\mathbf{X}_{k+1}) = \mathbf{0}$, 即

$$\nabla f(\mathbf{X}_k)^T + \mathbf{H}(\mathbf{X}_k) (\mathbf{X}_{k+1} - \mathbf{X}_k) = \mathbf{0}$$

当 $\mathbf{H}(\mathbf{X}_k)$ 可逆, 可解得

$$\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{H}(\mathbf{X}_k)^{-1} \nabla f(\mathbf{X}_k)$$

记 $\mathbf{P}_k = -\mathbf{H}(\mathbf{X}_k)^{-1} \nabla f(\mathbf{X}_k)$, 称为牛顿方向。

由上述分析可知, 对于正定二次函数 $f(\mathbf{X})$, 采用牛顿方向作搜索方向, 步长取为 1, 从任意点 \mathbf{X}_k 出发一步即可得极小点。对于一般函数 $f(\mathbf{X})$, 在一定条件下, 也可采用牛顿方向, 所形成的算法称为牛顿法。其一般步骤如下:

- ① 选取初始点 $\mathbf{X}_0, k := 0, \varepsilon > 0$;
- ② 对 \mathbf{X}_k , 若 $\|\nabla f(\mathbf{X}_k)\| < \varepsilon$, 则停止, 否则转③;
- ③ 取 $\mathbf{P}_k = -\mathbf{H}(\mathbf{X}_k)^{-1} \nabla f(\mathbf{X}_k)$, 由 $\min_{\lambda} f(\mathbf{X}_k + \lambda \mathbf{P}_k)$ 求出 $\lambda_k, \mathbf{X}_{k+1} = \mathbf{X}_k + \lambda_k \mathbf{P}_k$;
- ④ 计算 $\nabla f(\mathbf{X}_{k+1})$, 置 $k := k + 1$, 转②。

可以证明,当一维搜索是精确的,牛顿法为二阶收敛。尽管该算法具有收敛速度快的优点,但也存在一些不足。首先是它对函数 $f(\mathbf{X})$ 有较多的要求,如要求 $f(\mathbf{X})$ 的二阶导存在且其海赛阵是正定的,以保证逆阵存在且所构造的搜索方向是使 $f(\mathbf{X})$ 下降的;其次计算海赛阵的逆阵的工作量也可能较大。

因此,在牛顿法的基础上加以改进,设法构造一个矩阵 \mathbf{H}_k 代替牛顿方向中的 $\mathbf{H}(\mathbf{X}_k)^{-1}$,使 \mathbf{H}_k 满足:① \mathbf{H}_k 只用到 $f(\mathbf{X})$ 的一阶导信息;② \mathbf{H}_k 正定;③随着 k 的增大, \mathbf{H}_k 充分接近于 $\mathbf{H}(\mathbf{X}^*)^{-1}$ 。称 \mathbf{H}_k 为尺度阵。由于它每次都在变,故称以 \mathbf{H}_k 代替牛顿法中的 $\mathbf{H}(\mathbf{X}_k)^{-1}$ 的算法为变尺度法。

2. 拟牛顿法

拟牛顿法是尺度阵 \mathbf{H}_k 满足 $\mathbf{X}_k - \mathbf{X}_{k-1} = \mathbf{H}_k (\nabla f(\mathbf{X}_k) - \nabla f(\mathbf{X}_{k-1}))$ 的一类变尺度法。它一般是超线性收敛的。现介绍一种著名的拟牛顿方法——DFP法。它是由 Davidon(1959)提出的,后经 Fletcher 和 Powell(1963)修改及整理而成的。其一般步骤如下:

①选取初始点 \mathbf{X}_0 和初始对称正定阵 \mathbf{H} , (可取 $\mathbf{H}_0 = \mathbf{I}$) $k := 0, \epsilon > 0$;

②对 \mathbf{X}_k , 若 $\|\nabla f(\mathbf{X}_k)\| < \epsilon$, 则停止, 否则转③;

③取 $\mathbf{P}_k = -\mathbf{H}_k \nabla f(\mathbf{X}_k)$,

$$\mathbf{H}_k = \mathbf{H}_{k-1} + \frac{\Delta \mathbf{X}_{k-1} \cdot \Delta \mathbf{X}_{k-1}^T}{\Delta \mathbf{X}_{k-1}^T \cdot \Delta \mathbf{G}_{k-1}} - \frac{\mathbf{H}_{k-1} \Delta \mathbf{G}_{k-1} \Delta \mathbf{G}_{k-1}^T \mathbf{H}_{k-1}}{\Delta \mathbf{G}_{k-1}^T \mathbf{H}_{k-1} \Delta \mathbf{G}_{k-1}}$$

其中 $\Delta \mathbf{X}_{k-1} = \mathbf{X}_k - \mathbf{X}_{k-1}$, $\Delta \mathbf{G}_{k-1} = \nabla f(\mathbf{X}_k) - \nabla f(\mathbf{X}_{k-1})$, 沿 \mathbf{P}_k 进行一维搜索, 得 λ_k 和 $\mathbf{X}_{k+1} = \mathbf{X}_k + \lambda_k \mathbf{P}_k$;

④计算 $\nabla f(\mathbf{X}_{k+1})$, 置 $k := k + 1$, 转②。

可以证明, 当 $f(\mathbf{X})$ 为二次函数时, DFP 法是共轭方向法, 且具有二次终止性。

例 3.13 用 DFP 法计算 $f(\mathbf{X}) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1$ 的极小点, 取 $\mathbf{X}_0 = [-2 \ 4]^T$ 。

$$\text{解 } f(\mathbf{X}) = \frac{1}{2} [x_1 \ x_2] \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [-2 \ 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\text{取 } \mathbf{H}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\nabla f(\mathbf{X}) = [3x_1 - x_2 - 2 \ x_2 - x_1]^T, \nabla f(\mathbf{X}_0) = [-12 \ 6]^T$$

$$\mathbf{P}_0 = -\mathbf{H}_0 \nabla f(\mathbf{X}_0) = - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -12 \\ 6 \end{bmatrix} = \begin{bmatrix} 12 \\ -6 \end{bmatrix}, \mathbf{H}(\mathbf{X}_0) = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}$$

沿 \mathbf{P}_0 一维搜索, 由共轭方向最佳步长公式可得

$$\lambda_0 = \frac{-\nabla f(\mathbf{X}_0)^T \mathbf{P}_0}{\mathbf{P}_0^T \mathbf{H} \mathbf{P}_0} = \frac{[12 \ -6] \begin{bmatrix} 12 \\ -6 \end{bmatrix}}{[12 \ -6] \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 12 \\ -6 \end{bmatrix}} = \frac{5}{17}$$

$$\mathbf{X}_1 = \mathbf{X}_0 + \lambda_0 \mathbf{P}_0 = [-2 \ 4]^T + \frac{5}{17} [12 \ -6]^T = \left[\frac{26}{17} \ \frac{38}{17} \right]^T$$

$$\nabla f(\mathbf{X}_1) = \left[\frac{6}{17} \ \frac{12}{17} \right]^T$$

$$\begin{aligned}
\Delta \mathbf{X}_0 &= \mathbf{X}_1 - \mathbf{X}_0 = \begin{bmatrix} \frac{60}{17} & -\frac{30}{17} \end{bmatrix}^T, \Delta \mathbf{G}_0 = \nabla f(\mathbf{X}_1) - \nabla f(\mathbf{X}_0) = \begin{bmatrix} \frac{210}{17} & -\frac{90}{17} \end{bmatrix}^T \\
\mathbf{H}_1 &= \mathbf{H}_0 + \frac{\Delta \mathbf{X}_0 \Delta \mathbf{X}_0^T}{\Delta \mathbf{X}_0^T \Delta \mathbf{G}_0} - \frac{\mathbf{H}_0 \Delta \mathbf{G}_0 \Delta \mathbf{G}_0^T \mathbf{H}_0}{\Delta \mathbf{G}_0^T \mathbf{H}_0 \Delta \mathbf{G}_0} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\begin{bmatrix} \frac{60}{17} \\ -\frac{30}{17} \end{bmatrix} \begin{bmatrix} \frac{60}{17} & -\frac{30}{17} \end{bmatrix}}{\begin{bmatrix} \frac{60}{17} & -\frac{30}{17} \end{bmatrix} \begin{bmatrix} \frac{210}{17} \\ -\frac{90}{17} \end{bmatrix}} - \frac{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{210}{17} & -\frac{90}{17} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{210}{17} \\ -\frac{90}{17} \end{bmatrix}}{\begin{bmatrix} \frac{210}{17} & -\frac{90}{17} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{210}{17} \\ -\frac{90}{17} \end{bmatrix}} \\
&= \frac{1}{986} \begin{bmatrix} 385 & 241 \\ 241 & 891 \end{bmatrix} \\
\mathbf{P}_1 &= -\mathbf{H}_1 \nabla f(\mathbf{X}_1) = -\frac{1}{986} \begin{bmatrix} 385 & 241 \\ 241 & 891 \end{bmatrix} \begin{bmatrix} \frac{6}{17} \\ \frac{12}{17} \end{bmatrix} = -\begin{bmatrix} \frac{9}{29} \\ \frac{21}{29} \end{bmatrix}
\end{aligned}$$

沿 \mathbf{P}_1 一维搜索, 由共轭方向最佳步长公式

$$\lambda_1 = \frac{-\nabla f(\mathbf{X}_1)^T \mathbf{P}_1}{\mathbf{P}_1^T \mathbf{H} \mathbf{P}_1} = -\frac{\begin{bmatrix} \frac{6}{17} & \frac{12}{17} \end{bmatrix} \begin{bmatrix} \frac{9}{29} \\ \frac{21}{29} \end{bmatrix}}{\begin{bmatrix} \frac{9}{29} & \frac{21}{29} \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{9}{29} \\ \frac{21}{29} \end{bmatrix}} = \frac{29}{17}$$

$$\mathbf{X}_2 = \mathbf{X}_1 + \lambda_1 \mathbf{P}_1 = [1 \ 1]^T$$

$\nabla f(\mathbf{X}_2) = [0 \ 0]^T$, 故 $\mathbf{X}_2 = [1 \ 1]^T$ 是极小点。

3.3 约束极值问题

约束极值问题是更一般的非线性规划类型, 有广泛的实际背景。它的一般表达式为

$$\begin{aligned}
&\min f(\mathbf{X}) \\
&(\text{NLP}) \text{ s. t. } h_i(\mathbf{X}) = 0 \quad (i=1, \dots, m) \\
&\quad g_j(\mathbf{X}) \geq 0 \quad (j=1, \dots, l)
\end{aligned}$$

记其约束集为 D , 也可将其表示为

$$(\text{NLP}) \min_{\mathbf{X} \in D} f(\mathbf{X})$$

本节将首先介绍约束极值问题的最优性条件, 然后讨论一种特殊的约束极值问题——二次规划, 最后给出一般的约束极值问题的一类求解方法——罚函数法。

3.3.1 最优性条件

最优性条件即最优解的条件,包括必要条件、充分条件或充要条件。研究最优性条件对于算法设计及解的理论分析等具有重要意义。

1. 基本概念与性质

(1) 起作用约束

对于点 $X_0 \in D$, 在 X_0 点等于零的约束称为对 X_0 起作用的约束, 而在 X_0 点不等于零的约束称为对 X_0 不起作用的约束。

X_0 的起作用约束包括全体等式约束 $h_i(X) = 0 (i = 1, \dots, m)$ 和满足 $g_j(X_0) = 0 (j \in J \subset \{1, \dots, l\})$ 的不等式约束 $g_j(X) \geq 0 (j \in J)$ 。几何意义是 X_0 位于这些约束的边界上。 X_0 的不起作用约束包括使 $g_j(X_0) > 0 (j \in \bar{J})$ 的不等式约束 $g_j(X) \geq 0 (j \in \bar{J})$ 。几何意义是 X_0 位于这些约束的内部。

(2) 可行下降方向

方向 $P \in \mathbb{R}^n$ 在点 $X_0 \in D$ 称为是可行的, 是指存在正数 λ_0 , 使对一切 $\lambda \in [0, \lambda_0]$ 都有 $X_0 + \lambda P \in D$; P 在 X_0 是下降的, 是指对 $X_0 + \lambda P \in D$, 有

$$f(X_0 + \lambda P) - f(X_0) < 0$$

显然, 如果 X^* 是(NLP)的极小点, 则在 X^* 不存在可行下降方向。

(3) 可行下降方向的条件

由于等式约束 $h_i(X) = 0$ 可以化为用两个不等式约束 $h_i(X) \geq 0, -h_i(X) \geq 0$ 来表示, 所以有时在讨论约束条件时可只考虑不等式约束 $g_j(X) \geq 0$ 。

由可行下降方向的定义, 某方向 P 在 X_0 点若满足

$$\begin{cases} g_j(X_0 + \lambda P) \geq 0 & (j = 1, \dots, l) \\ f(X_0 + \lambda P) - f(X_0) < 0 \end{cases}$$

则 P 是 X_0 的可行下降方向。由泰勒公式

$$g_j(X_0 + \lambda P) \approx g_j(X_0) + \lambda \nabla g_j(X_0)^T P \quad (j = 1, \dots, l)$$

$$f(X_0 + \lambda P) - f(X_0) \approx \lambda \nabla f(X_0)^T P$$

其中 λ 是足够小的正数(可保证 $X_0 + \lambda P \in D$)。不难看出, 只要

$$\begin{cases} \nabla g_j(X_0)^T P > 0 \\ \nabla f(X_0)^T P < 0 \end{cases}$$

便可保证 P 是可行下降的。这可作为 P 可行下降的充分条件。其几何意义是, P 与 $\nabla g_j(X_0)$ 夹角锐角, 同时 P 与 $-\nabla f(X_0)$ 也夹角锐角。由此可知, 如果在 X_0 不存在可行下降方向, 则一定不存在 P 同时与 $\nabla g_j(X_0)$ 和 $-\nabla f(X_0)$ 夹角锐角。

2. 最优性条件(K-T 必要条件)

仍先考虑只含有不等式约束的(NLP)。设 $f(X)$ 与 $g_j(X)$ 均为连续可微函数, X^* 是(NLP)的最优解。

如果 X^* 是 D 的内点, 即约束对 X^* 是不起作用的, 则由无约束极值问题的极值必要条件, 有 $\nabla f(X^*) = 0$ 。如果 X^* 是某一个约束 $g_1(X) \geq 0$ 的边界点, 即 $g_1(X) \geq 0$ 对 X^* 起作

用,则必有 $\nabla g_1(\mathbf{X}^*)$ 与 $-\nabla f(\mathbf{X}^*)$ 反向(图 3.7),否则必有 \mathbf{P} 与二者同夹锐角,即 \mathbf{P} 是 \mathbf{X}^* 的可行下降方向,矛盾。故 $\nabla g_1(\mathbf{X}^*)$ 与 $\nabla f(\mathbf{X}^*)$ 同向,即存在 $\gamma_1 \geq 0$,使 $\nabla f(\mathbf{X}^*) = \gamma_1 \nabla g_1(\mathbf{X}^*)$ 。如果 \mathbf{X}^* 位于某两个约束 $g_1(\mathbf{X}) \geq 0$ 和 $g_2(\mathbf{X}) \geq 0$ 的交点,即 $g_1(\mathbf{X}) \geq 0$ 和 $g_2(\mathbf{X}) \geq 0$ 均对 \mathbf{X}^* 起作用,则必有 $\nabla f(\mathbf{X}^*)$ 位于 $\nabla g_1(\mathbf{X}^*)$ 和 $\nabla g_2(\mathbf{X}^*)$ 的夹角内部(图 3.8),否则也会有 \mathbf{P} 与 $-\nabla f(\mathbf{X}^*)$ 、 $\nabla g_1(\mathbf{X}^*)$ 和 $\nabla g_2(\mathbf{X}^*)$ 同夹锐角,即存在可行下降方向 \mathbf{P} ,矛盾。故有 γ_1 和 γ_2 使得下式成立:

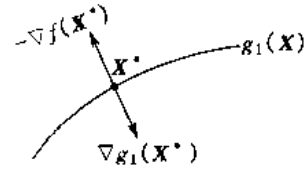


图 3.7

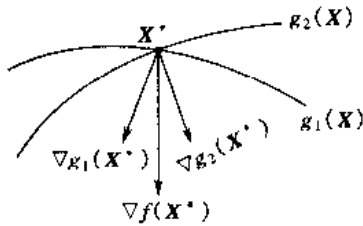


图 3.8

$$\nabla f(\mathbf{X}^*) = \gamma_1 \nabla g_1(\mathbf{X}^*) + \gamma_2 \nabla g_2(\mathbf{X}^*)$$

$$\gamma_1, \gamma_2 \geq 0$$

依此类推,若 \mathbf{X}^* 点起作用约束为 $g_j(\mathbf{X}) \geq 0, j \in J$,则有

$$\nabla f(\mathbf{X}^*) = \sum_{j \in J} \gamma_j \nabla g_j(\mathbf{X}^*)$$

$$\gamma_j \geq 0, j \in J$$

或表示为

$$\nabla f(\mathbf{X}^*) = \sum_{j=1}^l \gamma_j \nabla g_j(\mathbf{X}^*)$$

$$\gamma_j g_j(\mathbf{X}^*) = 0 \quad (j=1, \dots, l)$$

$$\gamma_j \geq 0 \quad (j=1, \dots, l)$$

其中的条件 $\gamma_j g_j(\mathbf{X}^*) = 0$ 是为了在梯度组合式中去掉不起作用约束,称为松紧条件。这时还要求 \mathbf{X}^* 点各起作用约束的梯度线性无关。

对于(NLP)中含有等式约束的情形,可将等式约束化为用两个不等式约束表示,并利用上述讨论结果,便得到下面的最优性必要条件,即著名 K-T(Kuhn Tucker)条件。

定理 3.1 (K-T 条件) 设 \mathbf{X}^* 是(NLP)的最优解,且在 \mathbf{X}^* 点各起作用约束的梯度 $\nabla h_i(\mathbf{X}^*) (i=1, \dots, m)$ 和 $\nabla g_j(\mathbf{X}^*) (j \in J)$ 线性无关,则存在向量 $\Lambda^* = [\lambda_1^* \dots \lambda_m^*]$ 和 $\Gamma^* = [\gamma_1^* \dots \gamma_l^*]$ 使下式成立:

$$\begin{cases} \nabla f(\mathbf{X}^*) = \sum_{i=1}^m \lambda_i^* \nabla h_i(\mathbf{X}^*) + \sum_{j=1}^l \gamma_j^* \nabla g_j(\mathbf{X}^*) \\ \gamma_j^* g_j(\mathbf{X}^*) = 0 (j=1, \dots, l) \\ \gamma_j^* \geq 0 (j=1, \dots, l) \end{cases}$$

K-T 条件中的 $\lambda_1^*, \dots, \lambda_m^*$ 和 $\gamma_1^*, \dots, \gamma_l^*$ 称为广义拉格朗日(Lagrange)乘子或 K-T 乘子。满足 K-T 条件表达式的可行点称做 K-T 点。可以证明, K-T 条件对于凸规划是充要的。因此,对于凸规划,可通过求 K-T 点来得到最优解。

例 3.14 利用 K-T 条件求解下面的非线性规划:

$$\min f(\mathbf{X}) = (x-3)^2$$

$$\text{s. t. } 0 \leq x \leq 5$$

解 原问题可写为

$$\min f(\mathbf{X}) = (x-3)^2$$

$$\begin{aligned} \text{s. t. } g_1(\mathbf{X}) &= x \geq 0 \\ g_2(\mathbf{X}) &= 5 - x \geq 0 \end{aligned}$$

计算目标和约束函数的海赛阵

$$\mathbf{H}_f(\mathbf{X}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0, \mathbf{H}_{g_1}(\mathbf{X}) = \mathbf{H}_{g_2}(\mathbf{X}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \leq 0$$

即 $f(\mathbf{X})$ 为凸函数, $g_1(\mathbf{X})$ 和 $g_2(\mathbf{X})$ 均为凹函数, 此问题是凸规划。

$$\nabla f(\mathbf{X}) = 2(x-3), \nabla g_1(\mathbf{X}) = 1, \nabla g_2(\mathbf{X}) = -1$$

K-T 条件表达式为

$$\begin{cases} 2(x-3) - \gamma_1 + \gamma_2 = 0 \\ \gamma_1 x = 0 \\ \gamma_2(5-x) = 0 \\ \gamma_1, \gamma_2 \geq 0 \end{cases}$$

为解此方程组, 可分几种情况考虑:

- ① $\gamma_1 \neq 0$ 且 $\gamma_2 \neq 0$, 无解;
- ② $\gamma_1 \neq 0$ 且 $\gamma_2 = 0$, 解得 $x = 0, \gamma_1 = -6$, 不是 K-T 点;
- ③ $\gamma_1 = 0$ 且 $\gamma_2 \neq 0$, 解得 $x = 5, \gamma_2 = -4$, 也不是 K-T 点;
- ④ $\gamma_1 = \gamma_2 = 0$, 解得 $x = 3$, 满足约束, 是 K-T 点。

由于此问题为凸规划, 故 $x^* = 3$ 是最优解, 相应最优值 $f(x^*) = 0$ 。

例 3.15 考虑下面的非线性规划:

$$\begin{aligned} \max f(\mathbf{X}) &= \ln(x_1 + 1) + x_2 \\ \text{s. t. } \begin{cases} 2x_1 + x_2 \leq 3 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned}$$

验证它为凸规划, 并用 K-T 条件求解。

解 原问题可写为

$$\begin{aligned} \min f(\mathbf{X}) &= -\ln(x_1 + 1) - x_2 \\ \text{s. t. } g_1(\mathbf{X}) &= 3 - 2x_1 - x_2 \geq 0 \\ g_2(\mathbf{X}) &= x_1 \geq 0 \\ g_3(\mathbf{X}) &= x_2 \geq 0 \end{aligned}$$

计算目标和约束函数的海赛阵

$$\mathbf{H}_f(\mathbf{X}) = \begin{bmatrix} \frac{1}{(x_1+1)^2} & 0 \\ 0 & 0 \end{bmatrix} \geq 0, \mathbf{H}_{g_1}(\mathbf{X}) = \mathbf{H}_{g_2}(\mathbf{X}) = \mathbf{H}_{g_3}(\mathbf{X}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \leq 0$$

故此问题是凸规划。

$$\begin{aligned} \nabla f(\mathbf{X}) &= \left[\frac{-1}{x_1+1} \quad -1 \right]^T, \nabla g_1(\mathbf{X}) = [-2 \quad -1]^T, \nabla g_2(\mathbf{X}) = [1 \quad 0]^T, \nabla g_3(\mathbf{X}) \\ &= [0 \quad 1]^T, \text{K-T 条件表达式为} \end{aligned}$$

$$\begin{cases} \frac{-1}{x_1+1} = -2\gamma_1 + \gamma_2 \\ -1 = -\gamma_1 + \gamma_3 \\ \gamma_1(3-2x_1-x_2) = 0 \\ \gamma_2 x_1 = 0 \\ \gamma_3 x_2 = 0 \\ \gamma_1, \gamma_2, \gamma_3 \geq 0 \end{cases}$$

若 $\gamma_1 = 0$, 则无解, 于是 $\gamma_1 \neq 0$, 有 $3 - 2x_1 - x_2 = 0$

令 $x_1 = 0, \gamma_3 = 0$, 则有

$$\begin{cases} 1 - 2\gamma_1 + \gamma_2 = 0 \\ 1 - \gamma_1 = 0 \\ 3 - x_2 = 0 \end{cases}$$

解得 $x_1 = 0, x_2 = 3, \lambda_1 = \lambda_2 = 1, \lambda_3 = 0$, 显然 $[0 \ 3]^T$ 是可行点, 从而是极小点。

3.3.2 二次规划

二次规划是一类简单的非线性规划, 其模型的一般表达式为

$$\begin{aligned} \min f(\mathbf{X}) &= \frac{1}{2} \mathbf{X}^T \mathbf{H} \mathbf{X} + \mathbf{C}^T \mathbf{X} \\ \text{s. t. } &\begin{cases} \mathbf{A} \mathbf{X} + \mathbf{b} \geq \mathbf{0} \\ \mathbf{X} \geq \mathbf{0} \end{cases} \end{aligned}$$

式中, \mathbf{A} 为 $m \times n$ 阶矩阵; \mathbf{b} 为 m 维非负向量; \mathbf{C} 为 n 维向量。由于二次规划的约束是线性的, 从而是凸集, 故当目标的二次项系数阵 $\mathbf{H} \geq \mathbf{0}$ (半正定) 时为凸规划, 可用 K-T 条件求解。

将模型的约束表示为

$$\begin{aligned} g_i(\mathbf{X}) &= [a_{i1} \ \cdots \ a_{in}] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + b_i \quad (i=1, \cdots, m) \\ g_{m+j}(\mathbf{X}) &= [0 \ \cdots \ 0 \ \underset{j}{1} \ 0 \ \cdots \ 0] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad (j=1, \cdots, n) \end{aligned}$$

则梯度

$$\begin{aligned} \nabla g_i(\mathbf{X}) &= [a_{i1} \ \cdots \ a_{in}]^T \quad (i=1, \cdots, m) \\ \nabla g_{m+j}(\mathbf{X}) &= [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T \quad (j=1, \cdots, n) \\ \nabla f(\mathbf{X}) &= \mathbf{H} \mathbf{X} + \mathbf{C} \end{aligned}$$

记相应于 $\nabla g_i(\mathbf{X})$ 的 K-T 乘子为 $\mathbf{Y}_1 = [y_1 \ \cdots \ y_m]^T$, 相应于 $\nabla g_{m+j}(\mathbf{X})$ 的 K-T 乘子为 $\mathbf{Y}_2 = [y_{m+1} \ \cdots \ y_{m+n}]^T$, 则由 K-T 条件及约束条件, 有

$$\begin{aligned} \mathbf{H} \mathbf{X} + \mathbf{C} &= \mathbf{A}^T \mathbf{Y}_1 + \mathbf{I} \mathbf{Y}_2 \\ \mathbf{Y}_1^T (\mathbf{A} \mathbf{X} + \mathbf{b}) &= 0 \\ \mathbf{Y}_2^T \mathbf{X} &= 0 \end{aligned}$$

$$AX + b \geq 0$$

$$X, Y_1, Y_2 \geq 0$$

在上面一组式子中,若先不考虑两个松紧条件,则其余均为线性表达式,即

$$-HX + A^T Y_1 + IY_2 = C$$

$$AX + b \geq 0$$

$$X, Y_1, Y_2 \geq 0$$

如果在上面第一式中增加人工变量 $Z = [z_1 \cdots z_n]^T$ (为使出现单位阵),并增加使人工变量之和极小化的目标函数(为使人工变量最后均为0),则可构造一个线性规划

$$\begin{aligned} \min z &= z_1 + \cdots + z_n \\ \text{s. t. } &\begin{cases} -HX + A^T Y_1 + IY_2 + \text{sgn}(C)Z = C \\ -AX + X_s = b \\ X, Y_1, Y_2, Z, X_s \geq 0 \end{cases} \end{aligned}$$

其中 $\text{sgn}(C)Z$ 表示 Z 中各分量的符号与 C 中相应分量的符号相同,是为了保证出现单位阵; X_s 是松弛变量。用单纯形法求解该线性规划的单纯形初表为表 3.3。

表 3.3

X_B	$B^{-1}b$	X	X_s	Y_1	Y_2	Z
Z	C	$-H$	0	A^T	I	$\text{sgn}(C)I$
X_s	b	$-A$	I	0	0	0

为使所求得的 X, Y_1, Y_2 还满足松紧条件

$$Y_1^T X_s = 0$$

$$Y_2^T X = 0$$

应在每次单纯形表的迭代中保证 Y_1 与 X_s 以及 Y_2 与 X 的相应分量不同时为基变量。

例 3.16 求解二次规划

$$\min f(X) = \frac{1}{2}(2x_1^2 + 2x_2^2) - 8x_1 - 10x_2$$

$$\text{s. t. } \begin{cases} 6 - 3x_1 - 2x_2 \geq 0 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

解 $H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, A = [-3 \quad -2], C = [-8 \quad -10]^T, b = 6$

$$Y_1 = y_1, Y_2 = [y_2 \quad y_3]^T$$

显然有 $H > 0$, 故此为一凸规划, 可用 K-T 条件求解。

$$\begin{aligned} &-HX + A^T Y_1 + IY_2 + \text{sgn}(C)Z \\ &= -\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -3 \\ -2 \end{bmatrix} y_1 + \begin{bmatrix} y_2 \\ y_3 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -8 \\ -10 \end{bmatrix} = C \\ &-AX + X_s = \begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + x_3 = 6 = b \end{aligned}$$

相应的线性规划为

$$\begin{aligned} \min z &= z_1 + z_2 \\ \text{s. t. } \begin{cases} 2x_1 + 3y_1 - y_2 + z_1 = 8 \\ 2x_2 + 2y_1 - y_3 + z_2 = 10 \\ 3x_1 + 2x_2 + x_3 = 6 \\ x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2 \geq 0 \end{cases} \end{aligned}$$

求解时还要满足松紧条件 $y_1 x_3 = 0$ 、 $y_2 x_1 = 0$ 、 $y_3 x_2 = 0$ 。

用单纯形法求解的过程如表 3.4。

表 3.4

C_B	X_B	$B^{-1}b$	0 x_1	0 x_2	0 x_3	0 y_1	0 y_2	0 y_3	1 z_1	1 z_2	θ
z_1	1	8	2			3	-1		1		4
z_2	1	10		2		2		-1		1	
x_3	0	6	[3]	2	1						2
			-2	-2		-5	1	1			
z_1	1	4		-4/3	-2/3	[3]	-1		1		4/3
z_2	1	10		2		2		-1		1	5
x_1	0	2	1	2/3	1/3						
				-2/3	1/3	-5	1	1			
y_1	0	4/3		-4/9	-2/9	1	-1/3		1/3		
z_2	1	22/3		26/9	4/9		2/3	-1	-2/3	1	33/13
x_1	0	2	1	2/3	1/3						3
				-26/9	-4/9		-2/3	1	2/3		
y_1	0	32/13				1					
x_2	0	33/13		1	2/13		3/13	-9/26	-3/13	9/26	
x_1	0	4/13	1		-4/39		-2/13	3/13			
									1	1	

求解的结果是

$$x_1 = \frac{4}{13}, x_2 = \frac{33}{13}, x_3 = 0, y_1 = \frac{32}{13}, y_2 = 0, y_3 = 0$$

在每一次单纯形表迭代中都考虑了松紧条件,如在第一张表中,按通常规则应让 y_1 进基,但考虑到与它构成松紧式的 x_3 已在基中,并且 y_1 进基也不会使 x_3 出基,故选择了 x_1 进基,从而所得解 $X^* = \left[\frac{4}{13} \quad \frac{33}{13} \right]^T$ 满足松紧条件,是原来二次规划的最优解。

3.3.3 罚函数法

罚函数法是一类比较简单实用的求解一般约束极值问题的方法。其基本思想是,将约束与目标组合在一起,化为无约束极值问题求解。罚函数法又可分为外点法与内点法,外点法是从可行域的外部逐步逼近最优解,而内点法是从可行域的内部逐步逼近最优解。本节只介绍外点法。

该方法的关键是基于(NLP)构造一个新的目标函数 $P(X, M)$, 称做罚函数。

$$P(X, M) = f(X) + M \sum_{j=1}^l [\min(0, g_j(X))]^2 + M \sum_{i=1}^m h_i^2(X)$$

式中 M 是一个充分大的正数, 称做罚因子。含有 M 的项称为罚项。显而易见, 当 X 是可行点时, 罚项为 0; 而当 X 不是可行点时, 罚项是很大的正数。对 $P(X, M)$ 求极小, 可采用无约束优化方法, 罚项能够保证使点 X 逐步趋近可行域。外点罚函数法的一般步骤如下。

- ① 选取初始罚因子 $M_1 > 0$, 令 $k := 1, \epsilon > 0$;
- ② 对于 M_k , 求解无约束极值问题 $\min P(X, M_k)$, 得 X_k ;
- ③ 若 X_k 满足可行性的精度要求

$$|h_i(X_k)| \leq \epsilon \quad (i = 1, \dots, m)$$

$$g_j(X_k) \geq -\epsilon \quad (j = 1, \dots, l)$$

则停止, X_k 即(NLP)的最优解, 否则, 取 $M_{k+1} > M_k$, 置 $k := k + 1$, 转②。

用罚函数法求解一些简单的非线性规划时, 可先求出 $P(X, M)$ 的驻点, 然后令 $M \rightarrow +\infty$ 得到 X^* 。

例 3.17 求解非线性规划

$$\min f(X) = \frac{1}{3}(x_1 + 1)^3 + x_2$$

$$\text{s. t. } \begin{cases} x_1 - 1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

解 构造罚函数

$$P(X, M) = \frac{1}{3}(x_1 + 1)^3 + x_2 + M[\min(0, x_1 - 1)]^2 + M[\min(0, x_2)]^2$$

为考察其驻点, 令

$$\frac{\partial P}{\partial x_1} = (x_1 + 1)^2 + 2M[\min(0, x_1 - 1)] = 0$$

$$\frac{\partial P}{\partial x_2} = 1 + 2M[\min(0, x_2)] = 0$$

当 $x_1 - 1 \geq 0$, 由上面的第一式有 $x_1 = -1$, 不可行; 当 $x_2 \geq 0$, 由第二式有 $1 = 0$, 矛盾, 说明可行域内无驻点。

对于可行域外的点(满足 $x_1 - 1 < 0, x_2 < 0$), 可解得驻点

$$\begin{cases} x_1 = \sqrt{M^2 + 4M} - M - 1 \\ x_2 = -\frac{1}{2M} \end{cases}$$

在该点, 海赛阵

$$H = \begin{bmatrix} 2(x_1 + 1) + 2M & 0 \\ 0 & 2M \end{bmatrix} > 0$$

说明该驻点是极小点。

令 $M \rightarrow +\infty$, 则 $\begin{cases} x_1 \rightarrow 1 \\ x_2 \rightarrow 0 \end{cases}$, 而 $[1 \ 0]^T$ 满足约束, 是最优解。

例 3.18 求解非线性规划

$$\begin{aligned} \min f(\mathbf{X}) &= x_1 + x_2 \\ \text{s. t. } &\begin{cases} -x_1^2 + x_2 \geq 0 \\ x_1 \geq 0 \end{cases} \end{aligned}$$

解 构造罚函数

$$P(\mathbf{X}, M) = x_1 + x_2 + M[\min(0, -x_1^2 + x_2)]^2 + M[\min(0, x_1)]^2$$

考察其驻点, 令

$$\frac{\partial P}{\partial x_1} = 1 + 2M[\min(0, -x_1^2 + x_2)](-2x_1) + 2M[\min(0, x_1)] = 0$$

$$\frac{\partial P}{\partial x_2} = 1 + 2M[\min(0, -x_1^2 + x_2)] = 0$$

对于可行域外的点, 有

$$\begin{cases} 1 + 2M(-x_1^2 + x_2)(-2x_1) + 2Mx_1 = 0 \\ 1 + 2M(-x_1^2 + x_2) = 0 \end{cases}$$

解得

$$\begin{cases} x_1 = -\frac{1}{2(1+M)} \\ x_2 = \frac{1}{4(1+M)^2} - \frac{1}{2M} \end{cases}$$

当 $M \rightarrow +\infty$ 时, $\begin{cases} x_1 \rightarrow 0 \\ x_2 \rightarrow 0 \end{cases}$, 而 $[0 \ 0]^T$ 是可行点, 故是最优解。

习 题

3.1 某投资者正在考虑 n 只股票的组合投资计划。设第 j 只股票的价格为 P_j , 每股平均收益为 μ_j , 第 i 只股票与第 j 只股票收益的协方差为 σ_{ij} ($i, j = 1, \dots, n$)。现有总资金为 B 。若投资者希望总平均收益与总平均风险(可表示为 $\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j$)之差极大化, 请写出投资计划模型。

3.2 分别用分数法和 0.618 法求函数

$$f(t) = t^2 - 6t + 2$$

在区间 $[0, 10]$ 上的极小点, 要求缩短后的区间长度不大于原区间长度的 3%。

3.3 分别用梯度法(迭代三次)和牛顿法求解下面的无约束极值问题

$$\min f(\mathbf{X}) = 2x_1^2 + x_2^2 + 2x_1 x_2 + x_1 - x_2$$

3.4 用共轭梯度法求二次函数

$$f(\mathbf{X}) = \frac{1}{2} \mathbf{X}^T \mathbf{A} \mathbf{X}$$

的极小点, 其中

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

3.5 用变尺度法求解

$$\min f(\mathbf{X}) = x_1^4 + (x_2 + 1)^2$$

取初始点 $\mathbf{X}_0 = [0 \ 0]^T$ 。

3.6 证明非线性规划问题

$$\min f(\mathbf{X})$$

$$\text{s. t. } \mathbf{X} \geq \mathbf{0}$$

的 K-T 条件表达式为

$$\begin{cases} \nabla f(\mathbf{X}^*) \geq \mathbf{0} \\ \nabla f(\mathbf{X}^*)^T \mathbf{X}^* = 0 \end{cases}$$

其中 \mathbf{X}^* 是其最优解,并说明这条件的几何意义。

3.7 由于某制造企业停产了某滞销产品而造成生产能力过剩,现该企业正在考虑利用这些生产能力进行 A、B、C 三种新产品的生产。设可利用的各种生产能力(机时)及每种新产品的单位消耗(机时)和单位产品成本如表 3.5 所示。

表 3.5

生 产 设 备 \ 产 品	A	B	C	可利用生产能力(机时)
铣 床	9	3	5	500
车 床	5	4	0	350
磨 床	3	0	2	150
单位成本	15	6	8	
销量为 x_i 时的价格	$10 + 40x_1^{-\frac{1}{3}}$	$5 + 15x_2^{-\frac{1}{4}}$	$10 + 20x_3^{-\frac{1}{2}}$	

据销售部门预测,产品 C 每周可销售 20 单位,产品 A 和 B 无限制。销售价格将随各产品的产量 x_1, x_2, x_3 而变动,估计的变动关系式如表最下行所示。该企业希望决定各产品每周各生产多少可使总利润最大。

- (1)建立使总利润最大的新产品生产计划模型;
- (2)验证该模型是否为凸规划;
- (3)说明该模型可以用什么方法求解。

3.8 用 K-T 条件求解

$$\max f(\mathbf{X}) = x_1 + x_2$$

$$\text{s. t. } \begin{cases} x_1^2 + x_2^2 \leq 1 \\ x_1, x_2 \geq 0 \end{cases}$$

3.9 求解二次规划

$$\min f(\mathbf{X}) = 2x_1^2 - 4x_1x_2 + 4x_2^2 - 6x_1 - 3x_2$$

$$\text{s. t. } \begin{cases} x_1 + x_2 \leq 3 \\ 4x_1 + x_2 \leq 9 \\ x_1, x_2 \geq 0 \end{cases}$$

3.10 用(外点)罚函数法求解

$$\min f(\mathbf{X}) = x_1^2 + x_2^2$$

$$\text{s. t. } x_2 = 1$$

第4章 多目标规划

多目标规划又称多目标最优化,其思想萌芽于1776年经济学中的效用理论。1896年,法国经济学家 V. Pareto(帕累托)首先在经济理论的研究中提出了多目标最优化问题。1951年,美国数理经济学家 T. C. Koopmans 从生产和分配的活动分析中考虑了多目标决策问题,并首次提出了多目标最优化问题解的概念,将其命名为“Pareto 解”(即有效解)。同年, H. W. Kuhn 和 A. W. Tucker 从数学规划论角度首次提出向量极值问题及有关概念。进入20世纪70年代,随着第一次国际多目标决策研讨会的召开及这方面专著问世,多目标决策问题的研究工作迅速、蓬勃地开展起来,到目前为止,已取得若干有价值的研究成果。

多目标决策问题不同于单目标问题。在单目标决策中,评价一个方案优劣的目标只有一个,而在多目标决策中,对方案 d_j 的评价需从 p 个目标上考虑,即由在这 p 个目标上的评价价值构成的向量去评价原有方案,即有

单目标问题 方案 $d_j \leftarrow$ 评价值 $f(d_j)$

多目标问题 方案 $d_j \leftarrow$ 评价值向量 $[f_1(d_j) \ f_2(d_j) \ \cdots \ f_p(d_j)]$

显而易见,对于单目标问题,只要根据方案间的目标评价值,总能分出孰优孰劣,因此是完全有序的。而多目标决策问题则不同,多目标决策问题的决策目标有多个,对方案的评价是基于一个向量,而向量一般是无法直接比较大小的,是半有序的。有鉴于此,将单目标问题最优解的概念平移到多目标问题中是不适宜的,这就有必要首先阐明多目标规划问题各种解的概念及其相互关系。

4.1 多目标规划模型及其解的概念

可以从一个简单例子引入多目标规划问题的一般形式。

例 4.1(喜糖问题) 设市场上有甲级糖及乙级糖,单价分别为10元/斤及5元/斤。今要筹办一桩婚事。“筹备小组”计划总花费不超过80元,糖的总斤数不少于10斤,甲级糖不少于5斤。问如何确定最佳的采购方案?

先确定此问题应满足的条件。不难看出,当设甲级糖数量为 x_1 ,乙级糖数量为 x_2 时,有

$$\begin{cases} 10x_1 + 5x_2 \leq 80 \\ x_1 + x_2 \geq 10 \\ x_1 \geq 5 \\ x_1, x_2 \geq 0 \end{cases}$$

在研究以什么为“最佳”的衡量指标时,“筹备小组”的成员们意见可能会发生分歧,原因是因为他们可能会提出各种各样的目标来。

如果要求总花费最少,即要求

$$f_1(x_1, x_2) = 10x_1 + 5x_2 \rightarrow \min$$

如果要求糖的总斤数最大,即要求

$$f_2(x_1, x_2) = x_1 + x_2 \rightarrow \max$$

如果要求甲级糖数量最大,即要求

$$f_3(x_1, x_2) = x_1 \rightarrow \max$$

多目标最优化模型与单目标最优化模型的区别主要是目标多于1个。在这些目标中,有的追求极大化,有的追求极小化,而极大化问题经过适当变换可化为极小化问题。因此,可以给出多目标最优化模型的一般形式如下:

决策变量 x_1, x_2, \dots, x_n

目标函数 $\min f_1(x_1, x_2, \dots, x_n)$

...

$$\min f_p(x_1, x_2, \dots, x_n)$$

$$\text{约束条件} \begin{cases} g_1(x_1, \dots, x_n) \geq 0 \\ \dots \\ g_m(x_1, \dots, x_n) \geq 0 \end{cases}$$

若记 $\mathbf{X} = [x_1 \ x_2 \ \dots \ x_n]^T$, $F(\mathbf{X}) = [f_1(\mathbf{X}) \ \dots \ f_p(\mathbf{X})]^T$, $R = \{\mathbf{X} | g_i(\mathbf{X}) \geq 0, i = 1, 2, \dots, m\}$ 表示约束集, $V\text{-}\min F(\mathbf{X})$ 表示对向量 $F(\mathbf{X})$ 中的各目标函数 $f_1(\mathbf{X}), \dots, f_p(\mathbf{X})$ 同等地进行极小化,则模型的一般形式为

$$(\text{VMP}) \begin{cases} V\text{-}\min (f_1(\mathbf{X}), \dots, f_p(\mathbf{X})) \\ g_i(\mathbf{X}) \geq 0 \quad (i = 1, 2, \dots, m) \end{cases}$$

简记为

$$(\text{VMP}) \begin{cases} V\text{-}\min F(\mathbf{X}) \\ \mathbf{X} \in R \quad R = \{\mathbf{X} | g_i(\mathbf{X}) \geq 0, i = 1, 2, \dots, m\} \end{cases}$$

求解如上的多目标最优化问题,即是在可行域 R 中寻找一个“最优解” \mathbf{X}^* 。如前所述,多目标问题中最优解的含义和单目标问题中最优解的含义是不相同的。对于(VMP)问题中任意两个可行解 $\mathbf{X}_1, \mathbf{X}_2 \in R$,根据其评价向量难以区分出优劣。因此,这里的“最优解” \mathbf{X}^* ,很多情况下只是基于某种准则下的满意解。

以下就给出多目标问题各种解的概念。为比较向量函数值的“大”“小”,首先需要引进向量空间中向量间的比较关系,即序的关系。

若记 $F^1 = [f_1^1 \ \dots \ f_p^1]$, $F^2 = [f_1^2 \ \dots \ f_p^2]$, 则

① $F^1 < F^2$ 意味着向量 F^1 的每个分量都要严格地小于向量 F^2 对应的分量,即对于 $i = 1, \dots, p$, 均有 $f_i^1 < f_i^2$;

② $F^1 = F^2$ 意味着向量 F^1 的每个分量都要等于相应 F^2 的每个分量,即对于 $i = 1, \dots, p$, 均有 $f_i^1 = f_i^2$;

③ $F^1 \leq F^2$ 意味着向量 F^1 的每个分量都要小于或等于 F^2 对应的分量,即对于 $i = 1, \dots, p$, 均有 $f_i^1 \leq f_i^2$;

④ $F^1 \leq F^2$ 意味着向量 F^1 的每个分量都要小于或等于向量 F^2 对应的分量,并且至少存在 F^1 的某一个分量要严格地小于向量 F^2 对应的分量,即对于 $i = 1, \dots, p$ 均有 $f_i^1 \leq f_i^2$, 并且至少存在某个 $i_0 (1 \leq i_0 \leq p)$, 有 $f_{i_0}^1 < f_{i_0}^2$ 。

可见, $F^1 \leq F^2$ 等价于 $F^1 \leq F^2$ 且 $F^1 \neq F^2$

由上述定义所确定的向量之间的序, 称做向量的自然序。

下面给出一般多目标极小化模型(VMP)各种解的概念。

定义 4.1 设 $X^* \in R$, 若对任意 $X \in R$, 均有 $F(X^*) \leq F(X)$, 则称 X^* 为问题(VMP)的绝对最优解。其全体记为 R_{ab}^* 。图 4.1 给出了两个目标、一维变量绝对最优解的例子。

图中的 x^* 既是 $f_1(x)$ 的极小点, 又是 $f_2(x)$ 的极小点, 因而是该问题的绝对最优解。但在大多数多目标最优化问题中, 这样的绝对最优解往往是不存在的。

为引出如下定义, 先考察一下什么是劣解。设 $X^0 \in R$, 若存在另一个可行解 $X^1 \in R$, 有 $F(X^1) \leq F(X^0)$, 说明对于 $i = 1, \dots, p$, 均有 $f_i(X^1) \leq f_i(X^0)$, 且至少存在某个 $i_0 (1 \leq i_0 \leq p)$ 有 $f_{i_0}(X^1) < f_{i_0}(X^0)$,

显然可行解 X^0 相对于 X^1 来说是劣的。这是因为 X^1 的 p 个目标值都不比 X^0 对应的 p 个目标值坏, 并且至少比有一个要好, 称 X^0 为劣解。显然, 在决策过程中, 任何劣解应该首先被去掉。为此, 可给出非劣解的定义如下。

定义 4.2 设 $X \in R$, 若不存在 $X \in R$, 使 $F(X) \leq F(\bar{X})$, 则称 \bar{X} 为问题的非劣解, 又称有效解或 Pareto 解。其全体记为 R_{eo}^* 。

显然, 对任意的 $X \in R$, 若它不是劣解就是有效解, 并且有效解仅仅是(VMP)问题的一个“不坏”的解, 或说可接受解。故若 \bar{X} 是(VMP)的有效解, 意味着在约束集 R 中已找不到一个 X , 使得对应的 $F(X) = [f_1(X) \ \dots \ f_p(X)]^T$ 中的每一个分目标值都不比 $F(\bar{X}) = [f_1(\bar{X}) \ \dots \ f_p(\bar{X})]^T$ 中的相应值更小, 并且 $F(X)$ 中至少有一个目标值要比 $F(\bar{X})$ 中的相应值为小。

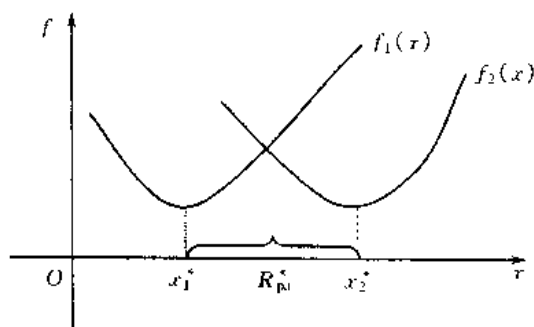


图 4.2

弱有效解。弱有效解的全体记为 R_{wp}^* 。

上述定义表明, 若 \bar{X} 是(VMP)的弱有效解, 则在约束集 R 中已找不到一个 X , 使得 $f_1(X), \dots, f_p(X)$ 中每一目标值都比 $f_1(\bar{X}), \dots, f_p(\bar{X})$ 中相应的值更小。

图 4.3 给出了弱有效解集合的例子。

显然, R_{wp}^* 中的任一点 \bar{x} 都为弱有效解, 因为找不到另一个 $x \in R$, 使 $F(x) < F(\bar{x})$ 。

在引进了“非劣”概念之后, 就可以对多目标规划作更确切的描述。准确地说, 多目标规划应该是从可行解中先选出有效解或弱有效解, 再根据某种准则, 从有效解中选择某种意义下的

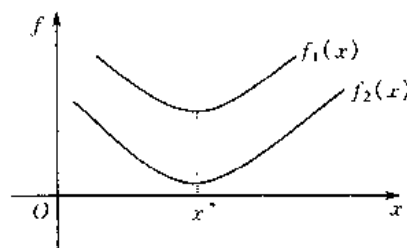


图 4.1

图 4.2 给出了两个目标、一维变量时有效解集合的例子。图中 x_1^* 与 x_2^* 之间的点 x 均为有效解 (x_1^* 与 x_2^* 之间的点不可比较), 因为找不到另一个可行解 x , 使 $F(x) \leq F(\bar{x})$ 成立。

把上述有效解定义中的条件稍微放宽, 可以给出下面的弱有效解的概念。

定义 4.3 设 $\bar{X} \in R$, 若不存在 $X \in R$, 使 $F(X) < F(\bar{X})$, 则称 \bar{X} 为问题(VMP)的

“最优解”,或说满意解或“好”的有效解。因为,从一般意义上说,现实问题中不存在这样一种方案,其实施结果按所有的判别标准各目标同时达到最佳。当然,能够找到这样的绝对最优解(方案)更好。

不妨观察下述两个目标极大化的决策问题。如图 4.4 所示,对于目标 f_1 ,方案 E 优于方案 D,而对于目标 f_2 ,方案 D 又优于方案 E,而方案 A、B、C 按现有准则均劣于方案 D、E。因此,在这 5 个方案中,方案 A、B、C 为劣解,方案 D、E 为有效解,则该问题的解只能按某种准则从方案 D 和 E 中选择其一。如若目标 f_1 比目标 f_2 重要得多,则可选方案 E。

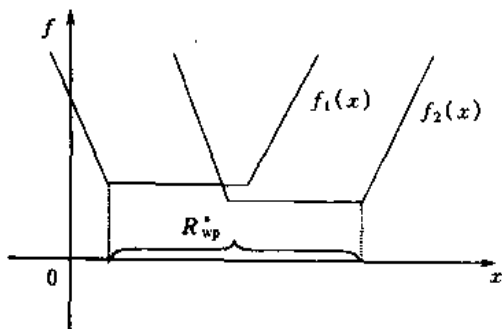


图 4.3

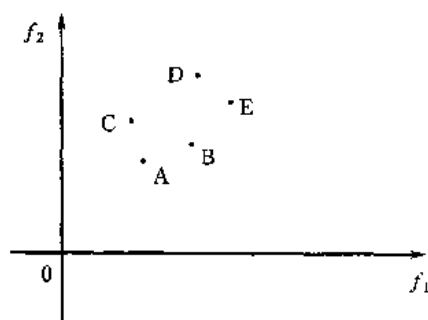


图 4.4

最后,不加证明地给出各种解之间的一些关系。

定理 4.1 $R_{ab}^* = \bigcap_{i=1}^p R_i^*$, 其中 R_i^* 为单目标 $f_i(X)$ 上最优点集合。

定理 4.2 $R_{pn}^* \subseteq R_{wp}^* \subseteq R$

定理 4.3 $R_i^* \subseteq R_{wp}^* \quad (i=1, \dots, p)$

定理 4.4 $R_{ab}^* \subseteq R_{pn}^*$

例 4.2 见图 4.5, $R_1^* = \{x_1\}$, $R_2^* = \{x_2\}$, 则 $R_{pn}^* = R_{wp}^* = [x_1, x_2]$

例 4.3 见图 4.6, R_1^* 、 R_2^* 、 R_{wp}^* 、 R_{pn}^* 的关系如图所示。

定理 4.5 设 $R_{ab}^* \neq \emptyset$, 则 ① $R_{pn}^* = \bigcap_{i=1}^p R_i^* = R_{ab}^*$; ② $R_{wp}^* = \bigcup_{i=1}^p R_i^*$ 。

文献 12 已经给出上述 5 个定理的证明,有兴趣的读者可以查看并试证。

本节讨论的解集之间的关系可用图 4.7 与图 4.8 表示。

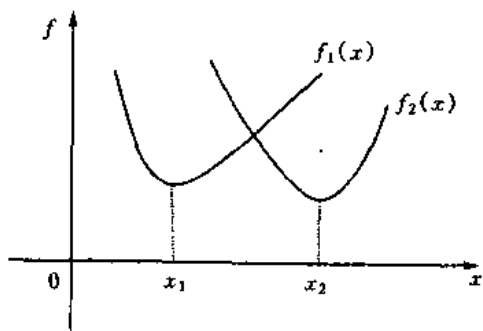


图 4.5

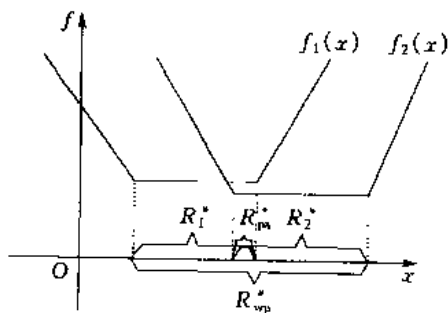


图 4.6

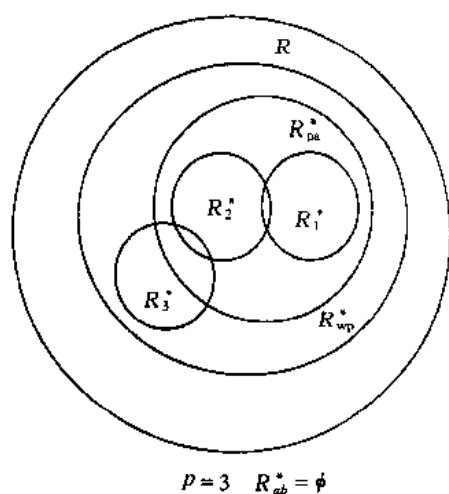


图 4.7

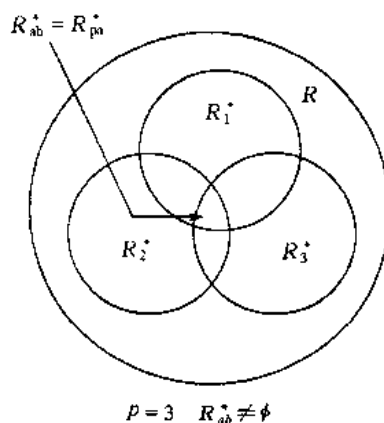


图 4.8

4.2 多目标规划的解法

求解多目标规划模型,就是要根据问题的特点和决策者的意图,选择适当算法,求得模型的有效解或弱有效解。本书介绍几类常用的多目标最优化问题的解法。

在实际问题中, p 个目标的量纲往往是不同的,所以事先需把每个目标规范化,然后再进行数学上的处理。例如,可以把带量纲的目标 $\bar{f}_i(\mathbf{X}) (i=1,2,\dots,p)$ 进行如下处理,令

$$f_i(\mathbf{X}) = \frac{\bar{f}_i(\mathbf{X})}{\bar{f}_i}, \text{ 其中 } \bar{f}_i = |\max_{\mathbf{X} \in R} \bar{f}_i(\mathbf{X})|$$

因此,以后总假定所讨论的多目标最优化模型已经过了无量纲化处理。

4.2.1 评价函数法

评价函数法的基本思想是利用一个复合函数将多目标问题转化为单目标问题求解。即对一个目标向量 $F(\mathbf{X}) = [f_1(\mathbf{X}) \cdots f_p(\mathbf{X})]$,构造一个把 p 个目标转化为一个数值目标的复合函数 $U[F(\mathbf{X})]$,作为问题的评价函数。即有:

$$\text{方案 } \mathbf{X} \leftarrow \text{评价向量 } F(\mathbf{X}) \leftarrow \text{评价函数 } U[F(\mathbf{X})]$$

这时,(VMP)问题转化为如下的单目标问题

$$\min_{\mathbf{X} \in R} U[F(\mathbf{X})]$$

求解该问题得最优的 \mathbf{X}^* ,则 \mathbf{X}^* 是该评价函数下的最优。进一步还可证明,利用评价函数法求出的单目标问题最优解 \mathbf{X}^* ,当该评价函数满足单调性时,一定是原多目标问题的有效解或弱有效解(证法略)。下面介绍7种常用的评价函数法。

1. 线性加权和法

线性加权和法就是取评价函数为各目标函数的线性加权和,即 $U(F) = \lambda_1 f_1 + \cdots + \lambda_p f_p$,其中 $\lambda_1, \dots, \lambda_p$ 为相应目标的权系数,其大小代表相应目标 f_i 在模型(VMP)中的重要程度。

线性加权和法的具体步骤如下:

①给出权系数,按各目标的重要性给出一组权系数 $\lambda_1, \dots, \lambda_p$, 其中 $\lambda_i \geq 0, \sum_{i=1}^p \lambda_i = 1$;

②极小化线性加权和函数,通过前述的线性加权和评价函数,将多目标问题转化为单目标问题

$$(P_\lambda) \min_{X \in R} \sum_{i=1}^p \lambda_i f_i(X)$$

求解 (P_λ) 得最优解 \bar{X} ;

③可以证明, \bar{X} 是(VMP)的弱有效解,即 $\bar{X} \in R_{wp}^*$, 进一步,若所有 $\lambda_i > 0$, 则 $\bar{X} \in R_{pw}^*$ 。

显然,线性加权和法的关键在于确定权系数。实际中,确定权系数的方法很多,如老手法、层次分析法、 α 法等。老手法及层次分析法将在本书后续内容中介绍。

α 方法是确定各目标重要程度的权系数的一种方法。它主要是根据 p 个目标的极小点信息,借助于引进一个辅助参数 α 的思想,通过求解一个 $p+1$ 阶线性方程组确定各目标的权系数。

先以两个目标为例。设两目标 $f_1(X)$ 及 $f_2(X)$ 皆为线性函数,都要求达到极小为最优,且约束集 R 为线性约束,根据评价函数法的基本思想,建立评价函数

$$U[F(X)] = \alpha_1 f_1(X) + \alpha_2 f_2(X)$$

问题归结为如何确定 α_1 及 α_2 的值。

这种方法的具体步骤如下。

①求解单目标优化问题

$$(\text{问题一}) \min_{X \in R} f_1(X) = f_1(X^1) \quad \text{记 } f_1^1 = f_1(X^1), f_2^1 = f_2(X^1)$$

$$(\text{问题二}) \min_{X \in R} f_2(X) = f_2(X^2) \quad \text{记 } f_2^2 = f_2(X^2), f_1^2 = f_1(X^2)$$

②由 $U[F(X^1)] = U[F(X^2)]$ 及 $\alpha_1 + \alpha_2 = 1$ 得下述方程组

$$\begin{cases} \alpha_1 f_1^1 + \alpha_2 f_2^1 = \alpha_1 f_1^2 + \alpha_2 f_2^2 \\ \alpha_1 + \alpha_2 = 1 \end{cases}$$

解上述方程组得

$$\alpha_1 = \frac{f_2^2 - f_2^1}{f_2^2 - f_2^1 + f_1^1 - f_1^2}$$

$$\alpha_2 = \frac{f_1^1 - f_1^2}{f_2^2 - f_2^1 + f_1^1 - f_1^2}$$

且

$$\frac{\alpha_1}{\alpha_2} = \frac{f_2^2 - f_2^1}{f_1^1 - f_1^2}$$

③求解 $\min_{X \in R} U[F(X)] = \min_{X \in R} [\alpha_1 f_1(X) + \alpha_2 f_2(X)]$ 得 X^* 。

以下对 α 方法进行几何解释。

如图 4.9 及 4.10 所示,可行域 R 中的点 X^1 及 X^2 经过映射在目标值空间 (f_1, f_2) 中分别对应 A 点及 B 点,即有

$$X^1 \rightarrow (f_1^1, f_2^1) A \text{ 点}$$

$$X^2 \rightarrow (f_1^2, f_2^2) B \text{ 点}$$

而且,在目标值空间中, $U[F(X)] = \alpha_1 f_1(X) + \alpha_2 f_2(X)$ 取不同数时,相当于一簇平行线,其斜率 $k = -\alpha_1/\alpha_2$,不妨记直线簇为 $\alpha_1 f_1 + \alpha_2 f_2 = c$ (c 为常数),而连结 A 、 B 的直线斜率也为 $k = -\alpha_1/\alpha_2$,则直线 AB 是该直线簇中的一条直线。显然,在同一条直线上 X^1 与 X^2 有相同的评价值,即有 $U[F(X^1)] = U[F(X^2)]$ 。进一步,要寻找评价值最小的 X ,可在目标值空间中平行移动直线。当直线与 C 点相切时,对应于 C 点的 X^* 之评价值为最优。

事实上,该方法可推至 p 个目标的情况。首先建立评价函数

$$U[F(X)] = \sum_{i=1}^p \alpha_i f_i(X)$$

然后求解一系列单目标问题

$$\min_{X \in R} f_i(X) = f_i(X^i) \text{ 得 } f_1^i, f_2^i, \dots, f_p^i$$

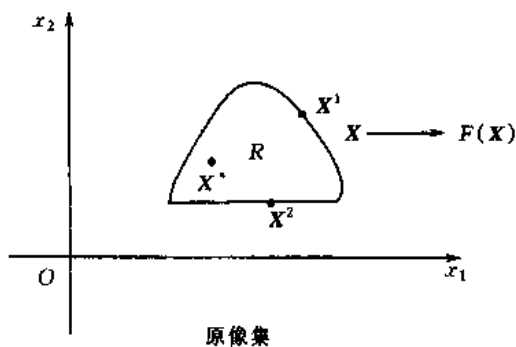


图 4.9

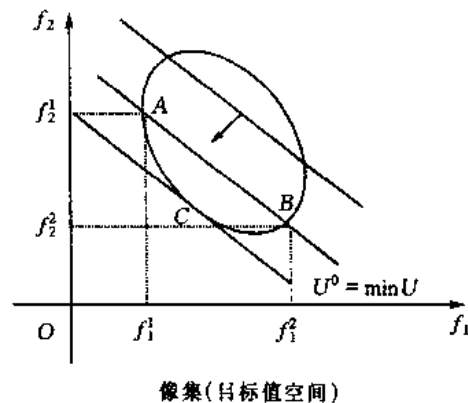


图 4.10

其中 f_j^i 的含义与二维情况相似 ($i, j = 1, 2, \dots, p$)。由于 $X^1, X^2, \dots, X^i, \dots, X^p$ 有相同的评价值,于是可建立下述方程组

$$\begin{cases} \alpha_1 f_1^1 + \alpha_2 f_2^1 + \dots + \alpha_p f_p^1 = c \\ \alpha_1 f_1^2 + \alpha_2 f_2^2 + \dots + \alpha_p f_p^2 = c \\ \vdots \\ \alpha_1 f_1^p + \alpha_2 f_2^p + \dots + \alpha_p f_p^p = c \\ \alpha_1 + \alpha_2 + \dots + \alpha_p = 1 \end{cases} \quad (c \text{ 为常数})$$

求解上述方程组得 $\alpha_1, \alpha_2, \dots, \alpha_p$ 。最后解 $\min_{X \in R} U[F(X)] = \min_{X \in R} [\alpha_1 f_1(X) + \dots + \alpha_p f_p(X)]$, 得 X^* 。

例 4.4 设有 $f_1(X) = 4x_1 + x_2 \rightarrow \min, f_2(X) = -3x_1 - 2x_2 \rightarrow \min$,

$R = \{X | 2x_1 + x_2 \leq 4, x_1 + x_2 \leq 3, x_1, x_2 \geq 0, X \in \mathbb{R}^2\}$, 试用 α 法求解。

解 先分别对 $f_1(X), f_2(X)$ 求得最优解, 它们是

$$f_1(X^1) = f_1(0, 0) = \min_{X \in \mathbb{R}^2} f_1(X) = f_1^1 = 0$$

$$f_2(X^2) = f_2(1, 2) = \min_{X \in \mathbb{R}^2} f_2(X) = f_2^2 = -7$$

则

$$f_2^1 = f_2(X^1) = f_2(0, 0) = 0$$

$$f_1^2 = f_1(\mathbf{X}^2) = f_1(1, 2) = 6$$

由此可得

$$\alpha_1 = \frac{f_2^2 - f_2^1}{f_2^2 - f_2^1 + f_1^1 - f_1^2} = \frac{-7}{-7-6} = \frac{7}{13}$$

$$\alpha_2 = \frac{6}{13}$$

$$U(\mathbf{X}) = \alpha_1 f_1(\mathbf{X}) + \alpha_2 f_2(\mathbf{X}) = \frac{7}{13} f_1(\mathbf{X}) + \frac{6}{13} f_2(\mathbf{X}) = \frac{10}{13} x_1 - \frac{5}{13} x_2$$

$$\text{易求得: } \min_{\mathbf{X} \in R} U(\mathbf{X}) = U(0, 3) = -\frac{15}{13}$$

2. 极大极小法

这种方法的基本出发点是考虑在对各个目标来说是最不利的情况下找出最有利的解。具体地说,对于多目标最优化模型(VMP),可以用各个目标 $f_i (i=1, \dots, p)$ 中的最大值作为评价函数的函数值,即取

$$U[F(\mathbf{X})] = \max_{1 \leq i \leq p} \{f_i(\mathbf{X})\}$$

为评价函数,通过上式把求解(VMP)归结为求解下述极小化问题:

$$\min_{\mathbf{X} \in R} U[F(\mathbf{X})] = \min_{\mathbf{X} \in R} \max_{1 \leq i \leq p} \{f_i(\mathbf{X})\}$$

并把它的最优解作为模型(VMP)的解。显然,这种方法是先对各目标函数作极大值选择之后,再在可行域上进行极小化,故称之为极大极小法。

图 4.11 给出两个目标、一维变量时由极大极小法所得解 \bar{x} 的图示。其中的实线为两个函数 $f_1(x)$ 和 $f_2(x)$ 中的最大值。

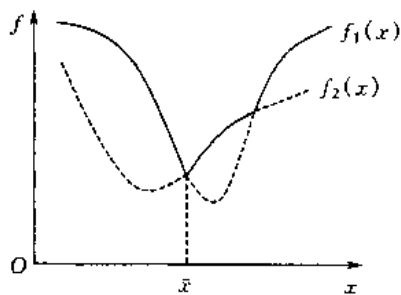


图 4.11

为了在评价函数中反映各目标的重要程度,一般考虑比上述极大极小化模型更一般的带权系数的情况。设对应于目标 f_i 的权系数为 $\lambda_i (i=1, 2, \dots, p)$, 则求解(VMP)相应地可归结为求解下述问题:

$$(P_\lambda) \min_{\mathbf{X} \in R} \max_{1 \leq i \leq p} \{\lambda_i f_i(\mathbf{X})\}$$

可以证明,当 $\lambda_i > 0 (i=1, 2, \dots, p)$ 时,求解 (P_λ) 所得的最优解为模型(VMP)的弱有效解。很明显,问题 (P_λ) 在实际求解时是不方便的。为此,引进一个变量,令

$$\lambda = \max_{1 \leq i \leq p} \{\lambda_i f_i(\mathbf{X})\}$$

这样,问题 (P_λ) 成为 $\min_{\mathbf{X} \in R} \lambda$, 且有

$$\lambda_i f_i(\mathbf{X}) \leq \lambda \quad (i=1, 2, \dots, p)$$

由此,问题 (P_λ) 便转化为如下的增加一个变量 λ 和 p 个约束条件的优化问题:

$$\begin{aligned} & \min \lambda \\ & \text{s. t. } \begin{cases} \mathbf{X} \in R \\ \lambda_i f_i(\mathbf{X}) \leq \lambda \quad (i=1, 2, \dots, p) \end{cases} \end{aligned}$$

根据以上讨论,可得出极大极小法的具体步骤如下:

①给出权系数,按各目标在(VMP)中的重要程度,确定相应的权系数 $\lambda_i (i=1, 2, \dots, p)$,

且 $\lambda_i > 0$;

②极小化辅助问题,即求解

$$\begin{aligned} & \min \lambda \\ & \text{s. t. } \begin{cases} \mathbf{X} \in R \\ \lambda_i f_i(\mathbf{X}) \leq \lambda \quad (i=1, 2, \dots, p) \end{cases} \end{aligned}$$

设所得最优解为 $[\mathbf{X}^* \quad \bar{\lambda}]^T$, 则 $\bar{\mathbf{X}}$ 即为模型(VMP)的弱有效解。

3. 平方和加权法

对于多目标问题(VMP), 设有 p 个规定值 f_1^*, \dots, f_p^* , 要求 p 个目标 $f_1(\mathbf{X}), \dots, f_p(\mathbf{X})$ 分别与规定的值相差尽量小, 若对其中不同值的要求相差程度又可不完全一样, 即有的要求重一些, 有的轻一些, 这时可采用下述评价函数:

$$U[F(\mathbf{X})] = \sum_{i=1}^p \lambda_i [f_i(\mathbf{X}) - f_i^*]^2, \quad \sum_{i=1}^p \lambda_i = 1 \text{ 且 } \lambda_i \geq 0$$

要求 $\min_{\mathbf{X} \in R} U[F(\mathbf{X})]$, 其中 λ_i 可按要求相差程度分别给出。其实, 平方和加权法是下面即将介绍的理想点法的变种。

4. 理想点法

理想点法的基本思想是认为 \mathbf{X} 的评价向量 $F(\mathbf{X}) = [f_1(\mathbf{X}) \quad f_2(\mathbf{X}) \quad \dots \quad f_p(\mathbf{X})]$ 越接近理想点越好。当然, 这种接近是指在某个模“ $\|\cdot\|$ ”的意义下评价向量与理想点之间的“距离”大小。所谓理想点一般指由各单目标最优值 f_1^*, \dots, f_p^* 组成的 p 维点 $\mathbf{F}^* = [f_1^* \quad \dots \quad f_p^*]$ 。具体地说, 理想点法就是在所定义的模的意义下, 以目标函数向量 $F(\mathbf{X})$ 与理想点之间的“距离” $U[F(\mathbf{X})] = \|F(\mathbf{X}) - \mathbf{F}^*\|$ 为评价函数, 把求解(VMP)的问题归结为求解下述数值极小化问题:

$$\min_{\mathbf{X} \in R} \|F(\mathbf{X}) - \mathbf{F}^*\|$$

下面以距离模评价函数为例说明这种方法的具体步骤。

①求理想点。求解 p 个单目标最优化问题

$$\min_{\mathbf{X} \in R} f_i(\mathbf{X}) = f_i(\mathbf{X}^i) = f_i^* \quad (i=1, 2, \dots, p)$$

得理想点 $\mathbf{F}^* = [f_1^* \quad \dots \quad f_p^*]$ 。

②检验理想点。若①中求出的各单目标最优解 \mathbf{X}^i 皆相同, 说明理想解点恰好是绝对最优点, 则输出绝对最优解 $\mathbf{X}^* = \mathbf{X}^i$, 求解完毕, 否则, 转③。

③作评价函数。取 q -模 $U[F(\mathbf{X})] = \|F(\mathbf{X}) - \mathbf{F}^*\| = \left[\sum_{i=1}^p (f_i^* - f_i(\mathbf{X}))^q \right]^{\frac{1}{q}}$ 为评价函数, q 的一般取值在 $[1, \infty)$, 通常取 $q=2$, 这时的评价函数即为欧氏空间中向量 $F(\mathbf{X})$ 与向量 \mathbf{F}^* 的距离, 其含义如图 4.12 所示。

④求解 $\min_{\mathbf{X} \in R} U[F(\mathbf{X})]$ 得 $\bar{\mathbf{X}}$ 。

可以验证, 上述的距离模评价函数关于 F 是严格增函数, 故由此所得的 $\bar{\mathbf{X}}$ 是模型(VMP)的有效解。

常用的评价函数还有:

①当 $q=1$ 时, $U[F(\mathbf{X})] = \sum_{i=1}^p |f_i^* - f_i(\mathbf{X})|$;

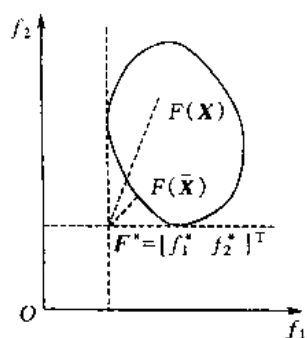


图 4.12

②当 $q = \infty$ 时, $U[F(X)] = \max_{1 \leq i \leq p} |f_i^* - f_i(X)|$ 。

理想点法其实是目标规划法的一类,即事先规定一些指标值,然后另设目标,看其接近这些规定值的程度,新设的目标有时也称超目标。可证,理想点法求出的解一定是非劣解,它在目标值空间中自然也就是有效点。

例 4.5 设 $f_1(X) = -3x_1 + 2x_2$, $f_2(X) = 4x_1 + 3x_2$ 都要实现最大,约束集为 $R = \{X | 2x_1 + 3x_2 \leq 18, 2x_1 + x_2 \leq 10, x_1 x_2 \geq 0, X \in R^2\}$,试用理想点法求解。

解 先分别求解两个单目标问题

$$\max_{X \in R^2} f_1(X) \text{ 及 } \max_{X \in R^2} f_2(X)$$

得最优解

$$X^{(1)} = (0, 6), X^{(2)} = (3, 4)$$

对应的目标值

$$f_1(X^{(1)}) = f_1(0, 6) = f_1^* = 12$$

$$f_2(X^{(2)}) = f_2(3, 4) = f_2^* = 24$$

故理想点

$$F^* = (f_1^*, f_2^*) = (12, 24)$$

取欧氏距离为评价函数,有

$$[(f_1(X) - f_1^*)^2 + (f_2(X) - f_2^*)^2]^{\frac{1}{2}}$$

可求得最优解 $X^* = [0.53 \quad 5.65]$, 对应的目标值 $f_1^* = 9.72, f_2^* = 19.06$ 。

5. 目标规划法

(1) 目标规划的数学模型

目标规划法由 A. Charnes 和 W. Cooper 首先提出。它是一种有着广泛应用的特殊多目标最优化模型。这类模型并不是去考虑对各个目标进行极小化或极大化,而是希望在约束条件的限制下,每一目标尽可能地接近于事先给定的目的值。它是求解多目标规划的一种常用方法,是理想点法的推广。

这种方法首先在每个目标 $f_i(X)$ 上预先确定一个希望达到的目的值 f_i^0 , 得目的值向量

$$F = [f_1^0 \quad f_2^0 \quad \cdots \quad f_p^0]$$

由此构造一个评价函数

$$\sum_{i=1}^p |f_i(X) - f_i^0|$$

然后把多目标决策问题转化为如下的单目标函数极值问题:

$$\min_{X \in R} \sum_{i=1}^p |f_i(X) - f_i^0|$$

它的含义是在确定的约束条件下,要求所有目标值尽可能接近希望达到的目的值,即要求偏差绝对值之和尽量小。为此,引入两类偏差变量 d_i^+, d_i^- ($i = 1, 2, \dots, p$), 有

$$\text{正偏差 } d_i^+ = \begin{cases} f_i(X) - f_i^0, & f_i(X) > f_i^0 \\ 0 & f_i(X) \leq f_i^0 \end{cases} \quad (i = 1, 2, \dots, p)$$

$$\text{负偏差 } d_i^- = \begin{cases} 0 & f_i(\mathbf{X}) \geq f_i^0 \\ f_i^0 - f_i(\mathbf{X}), & f_i(\mathbf{X}) < f_i^0 \end{cases} \quad (i=1,2,\dots,p)$$

d_i^+ 表示第 i 个目标所达到值超过预期目的值的部分; d_i^- 表示第 i 个目标值不足预期目的值的部分。则

$$d_i^+ + d_i^- = |f_i(\mathbf{X}) - f_i^0|$$

$$d_i^+ - d_i^- = f_i(\mathbf{X}) - f_i^0$$

$$d_i^+ \cdot d_i^- = 0 \quad (i=1,2,\dots,p) \text{ (目标达到值不可能既超过目的值又未到目的值)}$$

于是可以证明求解单目标决策问题

$$\min_{\mathbf{X} \in R} \sum_{i=1}^p |f_i(\mathbf{X}) - f_i^0|$$

等价于求解

$$\begin{cases} \min \sum_{i=1}^p (d_i^+ + d_i^-) \\ f_i(\mathbf{X}) - d_i^+ + d_i^- = f_i^0 \\ d_i^+ \cdot d_i^- = 0 \\ \mathbf{X} \in R, d_i^+, d_i^- \geq 0 \quad (i=1,2,\dots,p) \end{cases} \quad (4.1)$$

上述模型中含有偏差变量相乘的约束条件,这使得求解很不方便,考察去掉偏差变量相乘的约束条件,得到模型

$$\begin{cases} \min \sum_{i=1}^p (d_i^+ + d_i^-) \\ f_i(\mathbf{X}) - d_i^+ + d_i^- = f_i^0 \\ \mathbf{X} \in R \\ d_i^+ \geq 0, d_i^- \geq 0 \quad i=1,2,\dots,p \end{cases} \quad (4.2)$$

可以证明,模型式(4.2)的最优解也是模型式(4.1)的最优解,因而可将模型式(4.2)作为目标规划模型的一般形式。在上述模型中,约束集 R 称为绝对约束或说硬约束。其他可加入偏差变量的约束条件为目标约束或软约束。它是目标规划所特有的约束。

实际问题中,一个规划问题常常有若干个目标,但决策者在要求达到这些目标时,是有主次或轻重缓急的不同。凡要求首先达到的目标赋予优先因子 P_1 ,其次的目标赋予优先因子 P_2, \dots ,规定 $P_i \gg P_{i+1}$ 表示 P_i 比 P_{i+1} 有更大的优先权。即首先保证 P_1 级目标的实现,这时可不考虑次级目标;而 P_2 级目标是在实现 P_1 级目标的基础上考虑的,以此类推。若要区别具有相同优先因子的两个目标的差别,可分别赋予它们不同的权系数 ω_i 。这样,在上述目标规划模型一般形式的基础上,可建立如下的加权或分层目标规划模型

$$\begin{cases} \min \sum_{j=1}^l P_j \left(\sum_{i=1}^p \omega_{ji}^- d_i^- + \sum_{i=1}^p \omega_{ji}^+ d_i^+ \right) \\ f_i(\mathbf{X}) - d_i^+ + d_i^- = f_i^0 \\ \mathbf{X} \in R \\ d_i^+ \geq 0, d_i^- \geq 0 \quad (i=1,2,\dots,p) \end{cases} \quad (4.3)$$

最后考察目标规划的目标函数。由于在实际问题中,对各目标上预先设立的目标值的要

求不一样,这样目标函数(评价函数)的形式也不一样,归纳起来,基本形式有三种:

①希望各目标值与预期目的值之间超过或不足的偏差尽量小,即要求恰好达到目的值,这时的评价函数形式为 $\sum_{i=1}^p (d_i^+ + d_i^-)$;

②希望各目标值与预期目的值之间不足的偏差都尽量小,而超过的偏差不限,这时的评价函数形式为 $\sum_{i=1}^p d_i^-$;

③希望各目标值与预期目的值之间超过的偏差尽量小,即允许达不到目的值,这时的评价函数形式为 $\sum_{i=1}^p d_i^+$ 。

当然,也可以对某些目标要求极小化正偏差,而对另一些目标要求极小化负偏差。以下用例子说明目标规划问题具体的建模过程。

例 4.6 有一纺织厂生产尼龙布和棉布,平均生产能力都是 1 km/h,工厂生产能力为每周 80 h。根据市场预测,下周最大销售量为:尼龙布为 70 km,棉布 45 km。尼龙布利润为 2.5 元/m,棉布利润为 1.5 元/m。工厂领导的管理目标如下。 P_1 :保证职工正常上班,避免开工不足; P_2 :尽量达到最大销售量; P_3 :尽量减少加班时间,限制加班时间不得超过 10 h。

解 首先设决策变量 x_1, x_2 分别表示尼龙布和棉布的下周计划产量。

目标函数:

$$\text{工时 } f_1(\mathbf{X}) = x_1 + x_2 \rightarrow f_1^0 = 80$$

$$\text{尼龙布产量 } f_2(\mathbf{X}) = x_1 \rightarrow f_2^0 = 70$$

$$\text{棉布产量 } f_3(\mathbf{X}) = x_2 \rightarrow f_3^0 = 45$$

然后设相应于上述目标函数的正负偏差变量 d_i^+ 和 d_i^- ($i=1,2,3$)。由于决策者要求尼龙布和棉布的产量不同,即二者的重要性不一样,则可根据二者的单位利润比 $2.5:1.5=5:3$ 确定权系数。

于是根据目标函数优先级别的不同,可建立目标规划模型如下:

$$\begin{aligned} & \min [P_1 d_1^- + P_2 (5d_2^- + 3d_3^-) + P_3 (d_1^+)] \\ & \text{s. t. } \begin{cases} x_1 + x_2 + d_1^- - d_1^+ = 80 \\ x_1 + d_2^- - d_2^+ = 70 \\ x_2 + d_3^- - d_3^+ = 45 \\ d_1^+ \leq 10 \\ x_1, x_2, d_i^+, d_i^- \geq 0 \quad (i=1,2,3) \end{cases} \end{aligned}$$

(2)线性目标规划问题的求解方法

前面提出的目标规划模型式(4.2)已经化为单目标问题,特别当目标函数 $f_i(\mathbf{X})$ ($i=1, \dots, p$) 和约束 R 均为线性时,它就是一个线性规划问题。

以下通过举例说明在实际中常用的分层加权线性目标规划(约束及目标函数均为线性)问题的求解方法——单纯形法。

线性目标规划的数学模型结构与线性规划的数学模型结构没有本质区别,所以只要预先确定各目标重要程度的优先顺序,即可用单纯形法求解。但考虑到目标规划的数学模型的一

些特点,作以下规定:

①因目标规划问题的目标函数都是追求极小化,所以以检验数 $c_j - C_B B^{-1} P_j \geq 0 (j = 1, 2, \dots, n)$ 为最优准则;

②从每个检验数的整体来看检验数的正负首先决定于第一优先因子 P_1 的系数 α_{1j} 的正负,若 $\alpha_{1j} = 0$,这时此检验数的正负决定于第二优先因子 P_2 的系数 α_{2j} 的正负;下面可依次类推。

下面以实例说明解分层线性目标规划问题单纯形法的计算步骤。

例 4.7 某工厂欲生产 A 和 B 两种产品,因市场需要量大,销路不成问题。生产 A 和 B 产品所需设备工时、材料以及单位产品的利润如表 4.1 所示。

表 4.1

资 源 \ 产 品	A	B	限 量
设备工时(月/单位)	2	4	12
材料(百吨/单位)	3	3	12
利润(万元/单位)	4	3.2	

工厂在制定生产计划时,必须按重要程度考虑如下三个目标:

目标 P_1 要求生产这两种产品的利润最少要达到 12 万元

目标 P_2 要求 A 产品的产量尽可能为 B 产品产量的 1.5 倍

目标 P_3 为充分利用设备工时,必须使设备的空闲时间尽可能小

问工厂应如何决定产品 A 和 B 的产量?

解 首先根据题意建立目标规划模型。设 x_1 和 x_2 分别为 A、B 两产品的产量。根据预先提出的目标及资源限制,考虑问题的约束条件如下:

利润约束 $4x_1 + 3.2x_2 - d_1^+ + d_1^- = 12$

产品产量约束 $x_1 - 1.5x_2 - d_2^+ + d_2^- = 0$

设备工时约束 $2x_1 + 4x_2 + d_3^- = 12$

原材料约束 $3x_1 + 3x_2 \leq 12$

根据提出的目标要求,三个目标函数为 $P_1 d_1^-$ 、 $P_2 (d_2^+ + d_2^-)$ 、 $P_3 d_3^-$,使其和为最小,则可建立一个有目标优先级的线性目标规划模型:

$$\begin{aligned} & \min [P_1 d_1^- + P_2 (d_2^+ + d_2^-) + P_3 d_3^-] \\ & \text{s. t. } \begin{cases} 4x_1 + 3.2x_2 - d_1^+ + d_1^- = 12 \\ x_1 - 1.5x_2 - d_2^+ + d_2^- = 0 \\ 2x_1 + 4x_2 + d_3^- = 12 \\ 3x_1 + 3x_2 + x_3 = 12 \\ x_i \geq 0, d_i^- \geq 0, d_i^+ \geq 0 \quad (i=1,2,3) \end{cases} \end{aligned}$$

式中, d_1^+ 、 d_1^- 分别为对预定利润目标的超过量和不足量; d_2^+ 、 d_2^- 分别为 A 产品产量比 1.5 倍 B 产品产量的超过量和不足量; d_3^- 剩余的设备工时; x_3 松弛变量。

对该目标规划问题可使用线性规划中的单纯形法求解。由于预先确定了三个目标的优先次序

P_1, P_2, P_3 , 因此首先求解最高级 P_1 目标函数下的线性规划, 使 P_1 目标函数达到最优, 然后转向下一级目标, 继续求解 P_2 目标函数下的线性规划, 使 P_2 目标达到最大可能的满足, 再继续向下一级更低的目标 P_3 , 使其尽可能的满足。但在这个过程中, 低级目标上的满足不能影响所有高级目标上已获目标值数。因此, 在用单纯形法迭代的过程中, 确定进基变量的必要条件是这个变量在所有更高级目标函数下的检验数均为零。具体解法见表 4.2 的单纯形表。

表 4.2

	C_B	X_B	$B^{-1}b$	0 x_1	0 x_2	0 d_1^+	P_1 d_1^-	P_2 d_2^-	P_2 d_2^+	P_3 d_3^-	0 x_3
(一)	P_1	d_1^-	12	4	3.2	-1	1	0	0	0	0
	P_2	d_2^-	0	[1]	-1.5	0	0	-1	1	0	0
	P_3	d_3^-	12	2	4	0	0	0	0	1	0
	0	x_3	0	3	3	0	0	0	0	0	1
	检验数		P_1	-4↑	-3.2	1	0	0	0	0	0
			P_2	-1	1.5	0	0	2	0	0	0
			P_3	-2	-4	0	0	0	0	0	0
(二)	P_1	d_1^-	12	0	[9.2]	-1	1	4	-4	0	0
	0	x_1	0	1	-1.5	0	0	-1	1	0	0
	P_3	d_3^-	12	0	7	0	0	2	-2	1	0
	0	x_3	12	0	7.5	0	0	3	-3	0	1
	检验数		P_1	0	-9.2↑	1	0	-4	4	0	0
			P_2	0	0	0	0	1	1	0	0
			P_3	0	7	0	0	2	2	0	0
(三)	0	x_2	1.3	0	1	-0.11	0.11	0.43	-0.43	0	0
	0	x_1	1.95	1	0	-0.165	0.165	-0.355	0.355	0	0
	P_3	d_3^-	2.9	0	0	0.77	-0.77	-1	1	1	0
	0	x_3	2.25	0	0	[0.825]	0.825	-0.225	0.225	0	1
	检验数		P_1	0	0	0	1	0	0	0	0
			P_2	0	0	0	0	1	1	0	0
			P_3	0	0	-0.77↑	0.77	1	-1	0	0

本题的求解步骤可概括如下:

①根据模型建立初始单纯形表(表 4.2(一)), 由于上述问题中有三个目标函数 $P_1 d_1^-$ 、 $P_2(d_2^- + d_2^+)$ 、 $P_3 d_3^-$, 所以表中也有三行检验数行, 分别为基本可行解在三个目标函数下的检验数;

②以最高级目标 P_1 为目标函数, 在单纯形表上进行迭代计算, 直至 P_1 下检验数均为非负, 即表示最高级目标函数 P_1 已达最优值, 见表 4.2(一)、(二)、(三);

③转入第二级目标 P_2 , 由表 4.2(三)中知, P_2 下的检验数也均为非负, 表示当前之基本可行解对于 P_2 目标函数也是最优, 顺次转入对第三级目标函数的迭代;

④在表 4.2(三)中, P_3 下检验数行绝对值最大的负检验数为 -1, 但其相应于 P_2 检验数行中的检验数为 1, 因此变量 d_2^- 不能选为进基变量, 而变量 d_1^+ 在 P_3 下检验数为 -0.77, 而在 P_1, P_2 下检验数均为零, 因此选择 d_1^+ 进基能改善 P_3 目标值, 而不影响第一级和第二级目标函数值;

⑤按此规则继续进行迭代直至不能进行为止(表 4.3), 最优解为 $x_1 = 2.4, x_2 = 1.6, d_1^- = 2.72, d_3^- = 0.8, d_2^- = d_2^+ = 0, x_3 = 0$ 。

表 4.3

C_B	X_B	$B^{-1}b$	0 x_1	0 x_2	0 d_1^+	P_1 d_1	P_2 d_2^+	P_2 d_2	P_3 d_3	0 x_3
0	x_2	1.6	0	1	0	0	0.4	-0.4	0	0.13
0	x_1	2.4	1	0	0	0	-0.4	0.4	0	0.2
P_3	d_3^-	0.8	0	0	0	0	-0.79	0.79	1	-0.93
0	d_1^+	2.72	0	0	1	1	-0.273	0.273	0	1.21
检验数	P_1	0	0	0	0	1	0	0	0	0
	P_2	0	0	0	0	0	1	1	0	0
	P_3	0	0	0	0	0	0.79	-0.79	0	0.93

事实上,线性规划问题可视为线性目标规划问题的一种特殊情况,因为对于下述线性规划问题:

$$\begin{aligned}
 & \min \sum_{j=1}^n c_j x_j \\
 \text{(LP)} \quad & \begin{cases} \sum_{j=1}^n a_{ij} x_j = b_i & (i=1, \dots, m) \\ x_j \geq 0 & (j=1, \dots, n) \end{cases}
 \end{aligned}$$

若将目标函数的下界取为 b_{m+1} , 则上述(LP)问题可转化为下述线性目标规划问题(GLP):

$$\begin{aligned}
 & \min [P_1(D^-, D^+), P_2(D^-, D^+)] \\
 \text{(GLP)} \quad & \begin{cases} \sum_{j=1}^n a_{ij} x_j + d_i^- - d_i^+ = b_i & (i=1, \dots, m) \\ \sum_{j=1}^n c_j x_j + d_{m+1}^- - d_{m+1}^+ = b_{m+1} & x_j \geq 0 (j=1, \dots, n) \end{cases}
 \end{aligned}$$

式中 $D^+ = [d_1^+ \ d_2^+ \ \dots \ d_{m+1}^+]$, $D^- = [d_1^- \ d_2^- \ \dots \ d_{m+1}^-]$ 。

线性规划只解决单目标问题,而目标规划部分地消除了这种局限。目标规划能处理带有多个目标的多目标决策问题,而且目标函数可以由具有不同度量单位的偏差变量构成。与线性规划相比,目标规划的真正价值在于按照决策者的目标优先结构,去求解有矛盾目标的多目标决策问题。

6. 分目标乘除法

在 p 个目标 $f_1(X), \dots, f_p(X)$ 中,其中 k 个目标 $f_1(X), \dots, f_k(X)$ 要求实现最小,其余的目标 $f_{k+1}(X), \dots, f_p(X)$ 要求实现最大,这时的多目标最优化模型成为混合最优化模型,记为(VHP)。分目标乘除法以及后面即将介绍的功效系数法都是求解多目标混合最优化模型的方法。下面首先介绍分目标乘除法的原理与步骤。

分目标乘除法的主要特点是将模型中的各分目标函数进行相乘和相除处理后,在可行域上求解。已经知道,为使 $f_1(X), \dots, f_k(X)$ 极小化,可以考虑它们的乘积极小化,而要使 $f_{k+1}(X), \dots, f_p(X)$ 极大化,又可以考虑让这些目标函数乘积的倒数极小化。因此,求解(VHP)问题便可以归为求解

$$\min_{X \in R} \frac{f_1(X) \cdots f_k(X)}{f_{k+1}(X) \cdots f_p(X)}$$

由上述数值极小化问题所得的最优解,显然是使位于分子的各目标取尽可能小的值,且使

位于分母的各目标取尽可能大的值的解。

以上所述方法实际上是对模型构造了下述评价函数：

$$U[F(\mathbf{X})] = \frac{f_1(\mathbf{X})f_2(\mathbf{X})\cdots f_k(\mathbf{X})}{f_{k+1}(\mathbf{X})\cdots f_p(\mathbf{X})}$$

应指明的是,为使 $U[F(\mathbf{X})]$ 有意义,在使用分目标乘除法求解时,一般要求各目标函数在可行域 R 上均取正值。

分目标乘除法的具体步骤如下。

第1步:分目标正值化。检查 $f_1(\mathbf{X}), \dots, f_p(\mathbf{X})$ 在 R 上是否均取大于零的值。若是,则进行下一步,否则,选取某实数 $M > 0$,使

$$f_i(\mathbf{X}) + M > 0 \quad \mathbf{X} \in R \quad (i=1, \dots, p)$$

并令 $\bar{f}_i(\mathbf{X}) = f_i(\mathbf{X}) + M (i=1, 2, \dots, p)$, 进行第二步。

第2步:极小化分目标乘除问题。求解

$$\min_{\mathbf{X} \in R} \frac{f_1(\mathbf{X}) \cdots f_k(\mathbf{X})}{\bar{f}_{k+1}(\mathbf{X}) \cdots \bar{f}_p(\mathbf{X})}$$

设得最优解为 $\bar{\mathbf{X}}$, 则可以证明 $\bar{\mathbf{X}}$ 为(VHP)的有效解。

分目标乘除法有简单的经济学含义。以两目标的最优化模型(VHP)为例,设 $f_1(\mathbf{X})$ 表示投资, $f_2(\mathbf{X})$ 表示收益,则乘除分目标函数 $f_1(\mathbf{X})/f_2(\mathbf{X})$ 表示单位收益的投资,其倒数表示单位投资的收益即投资利润率。就此,分目标乘除法的意义是:极小化单位收益的投资。

例 4.8 某投资开发公司拥有总资金 100 万元,今有 4 个项目可供选择投资。投资第 i 个项目所用资金 a_i 及预计可得收益 b_i 为: $a_1 = 40$ (万元), $a_2 = 50$ (万元), $a_3 = 35$ (万元), $a_4 = 40$ (万元); $b_1 = 30$ (万元), $b_2 = 40$ (万元), $b_3 = 25$ (万元), $b_4 = 35$ (万元)。试用分目标乘除法决策投资方案。

解 设 $x_i = \begin{cases} 1, & \text{决定投资第 } i \text{ 个项目} \\ 0, & \text{决定不投资第 } i \text{ 个项目} \end{cases} \quad (i=1, 2, 3, 4)$

首先,由于 x_i 不全为零,故分目标函数 $f_1(\mathbf{X}) = 40x_1 + 50x_2 + 35x_3 + 40x_4$ 和 $f_2(\mathbf{X}) = 30x_1 + 40x_2 + 25x_3 + 35x_4$ 在优化过程中在 R 上取值大于零。

由此,可构造下述极小化分目标乘除问题:

$$\min_{\mathbf{X} \in R} \frac{40x_1 + 50x_2 + 35x_3 + 40x_4}{30x_1 + 40x_2 + 25x_3 + 35x_4}$$

求解上述分式规划问题,得最优解 $\bar{\mathbf{X}} = [0 \quad 1 \quad 0 \quad 1]^T$ 。即该公司应该投资第 2 和第 4 个项目。且总投资额 $f_1(\bar{\mathbf{X}}) = 50 + 40 = 90$ 万元,总收益 $f_2(\bar{\mathbf{X}}) = 40 + 35 = 75$ 万元。

7. 功效系数法——几何平均法

设 p 个目标 $f_1(\mathbf{X}), \dots, f_p(\mathbf{X})$, 其中 k_1 个目标要求实现最大, k_2 个目标要求实现最小,其余的目标是过大不行,过小也不行。功效系数法是一种无量纲化方法,基本做法就是分别给这些目标 $f_i(\mathbf{X})$ 以一定的功效系数(评分) d_i 。 d_i 是在 $[0, 1]$ 之间的某一数。当目标最满意达到时,取 $d_i = 1$; 当最差时,取 $d_i = 0$ 。通常把描述 d_i 与 $f_i(\mathbf{X})$ 的关系式称为功效函数,表示为 $d_i = d_i[f_i(\mathbf{X})]$ 。显然,对于不同类型目标应选用不同类型的功效函数。

功效函数的构造方法很多,有线性插值法、指数法、折线法等。本书只简要介绍前两种方

法。

(1) 线性插值法

图 4.13 给出不同类型目标线性插值法的功效函数图示。

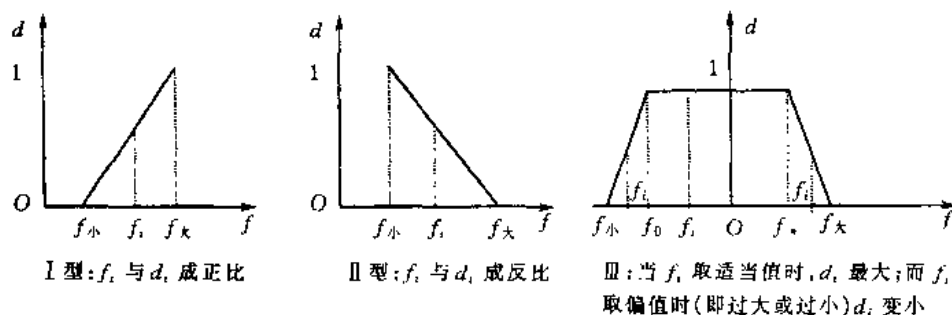


图 4.13

利用线性插值法得:

$$\text{I 型功效函数 } d_i = \frac{f_i - f_{\text{小}}}{f_{\text{大}} - f_{\text{小}}} \quad (\text{目标最优值为 } f_{\text{大}})$$

$$\text{II 型功效函数 } d_i = \frac{f_{\text{大}} - f_i}{f_{\text{大}} - f_{\text{小}}} \quad (\text{目标最优值为 } f_{\text{小}})$$

$$\text{III 型功效函数 } d_i = \begin{cases} 1 - (f_0 - f_i) / \Delta f^1, & \text{若 } f_i < f_0 \\ 1, & \text{若 } f_0 \leq f_i \leq f_* \\ 1 - (f_i - f_*) / \Delta f^2, & \text{若 } f_i > f_* \end{cases}$$

其中 $\Delta f^1 = \Delta f^2 = \max\{(f_{\text{大}} - f_*), (f_0 - f_{\text{小}})\}$, 目标最优值为区间 $[f_0, f_*]$ 。

在 III 型功效函数中, 当 $f_0 = f_*$ 时, 功效函数为

$$d_i = 1 - \frac{|f_i - f_0|}{\max_i |f_i - f_0|} \quad (\text{目标最优值为固定值 } f_0)$$

(2) 指数法

图 4.14 给出不同类型目标指数法的功效函数图示。图中 I 型、II 型、III 型功效函数的含义同线性插值法。

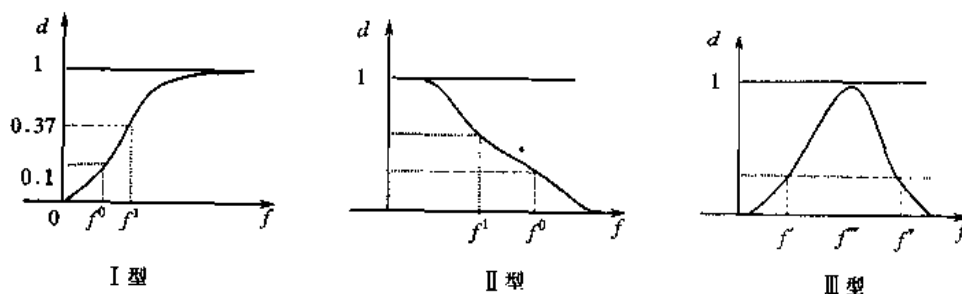


图 4.14

首先用指数法构造 I 型功效函数, 可设其表达式为

$$d = e^{-(e^{(b_0 + b_1 f)})}$$

其中 b_0, b_1 可这样确定:

当 f 达到某一刚合格值 f^1 时,取 $d^1 = e^{-1} \approx 0.3679$;

当 f 达到某一不合格值 f^0 时,取 $d^0 = e^{-e} \approx 0.06598$ 。

将上述要求代入上式即有

$$d^1 = e^{-1} = e^{-(b_0 + b_1 f^1)}$$

$$d^0 = e^{-e} = e^{-(b_0 + b_1 f^0)}$$

由此可得

$$\begin{cases} b_0 + b_1 f^1 = 0 \\ b_0 + b_1 f^0 = -1 \end{cases}$$

解之得

$$b_0 = \frac{f^1}{f^0 - f^1} \quad b_1 = \frac{-1}{f^0 - f^1}$$

即得 I 型功效函数

$$d = e^{-e \left(\frac{f - f^1}{f^0 - f^1} \right)}$$

同样对 II 型功效函数,可取为

$$d = 1 - e^{-e \left(\frac{f - f^1}{f^0 - f^1} \right)}$$

对于 III 型功效函数,取

$$d = e^{-(|Y|)^n}$$

式中
$$Y = \frac{2f - (f' + f'')}{f' - f''}$$

这样,当 $f = f'$ 或 $f = f''$ 时, $Y = \pm 1$, $d = e^{-1}$, 为刚好可接受的值;当 $f = \frac{f' + f''}{2}$ 时, $Y = 0$, $d = 1$, 即 f 达到比较适当的值。为了确定 n , 可另取一个 f''' , 使其与某一个适当的 d 值相对应, 这时可给出

$$n = \frac{\ln \ln \frac{1}{d}}{\ln |Y|}$$

如取 $f = f''' = \frac{2f' + f''}{3}$, 此时 $Y = \frac{1}{3}$, 使其与 $d = e^{-\frac{1}{2}}$ 相对应, 则

$$n = \frac{\ln \frac{1}{2}}{\ln \frac{1}{3}} \approx 0.6309$$

即

$$d = e^{-|Y|^{0.6309}}, Y = \frac{2f - (f' + f'')}{f' - f''}$$

有了功效函数后, 对每个目标都可对应为相应的功效函数, 目标值可转换为功效系数。这样, 可用各目标功效系数的几何平均值作为评价函数, 即有

$$U[F(X)] = \left[\prod_{i=1}^P d_i(f_i(X)) \right]^{\frac{1}{P}}$$

然后,极大化几何平均功效问题,即求解

$$\max_{X \in R} U[F(X)]$$

得最优解 \bar{X} 。

显然, $U[F(X)]$ 的值等于 1 是最满意的,等于零是最差的。这样定义评价函数有一个好处,一个方案中只要有一个目标值太差,整个评价值便会下降,该方案很难被选中。但这样有时也会造成决策判断上的偏差。如被评方案中有一个相对不太重要的目标的功效系数为零,而其他目标对应的功效系数很高,则这时该方案的评价值为零,该方案不会被选中。鉴于这种情况,应构造新的评价函数,或寻求新的无量纲化方法(如使功效系数的最小值不为零)。对此,已有若干研究成果,限于篇幅,本书不做详细介绍。

4.2.2 目标排序法

目标排序法的思想是把目标按重要性排序。设给出的重要性序列为 $f_1(X)$ 、 $f_2(X)$ 、 \dots 、 $f_p(X)$,然后按这种排序逐步进行一系列单目标优化,最后求出满意解。

1. 序列最优化法

首先对第一个目标求最优,并找出所有最优解的集合,记为 R_1 ,然后在 R_1 内求第二个目标的最优解,记这时的最优解集合为 R_2 ,如此等等,一直到求出第 p 个目标的“最优解” X_p 。基本步骤如下:

$$\begin{aligned} \min_{X \in R} f_1(X) &= f_1^*(X_1) && \text{得最优解集 } R_1 \\ \min_{X \in R_1 \subset R} f_2(X) &= f_2^*(X_2) && \text{得最优解集 } R_2 \\ &\vdots && \vdots \\ \min_{X \in R_{p-2} \subseteq R_{p-1}} f_{p-1}(X) &= f_{p-1}^*(X_{p-1}) && \text{得最优解集 } R_{p-1} \\ \min_{X \in R_{p-1} \subseteq R_{p-2}} f_p(X) &= f_p^*(X_p) && \text{得 } X_p \text{ 为“最优解”} \end{aligned}$$

使用序列最优化方法求得的解 X_p 一定是原多目标规划的非劣解,是分层序列意义下的最优解,然而这种方法可能由于某一步上最优解是惟一时,以后的目标优化就不能进行下去了。

2. 宽容意义下的序列优化方法

为了避免 R_i 为空集或某一步上最优解是惟一时不可继续进行的困难,可采用放宽条件下进行分层序列优化,即预先给定 $\epsilon_1 > 0$ 、 $\epsilon_2 > 0$ 、 \dots 、 $\epsilon_{p-1} > 0$ 作为宽容值。这些 ϵ_i 是根据各目标不同要求预先给出,它是相对于这目标最优值之允许偏差。每一分层单目标优化是在相对最优解有宽容量 ϵ_i 的约束集中进行的。基本步骤如下:

$$\begin{aligned} \min_{X \in R} f_1(X) &= f_1^* \\ \min_{X \in R_1} f_2(X) &= f_2^*, R_1 = \{X | f_1(X) < f_1^* + \epsilon_1, X \in R\} \\ \min_{X \in R_2} f_3(X) &= f_3^*, R_2 = \{X | f_2(X) < f_2^* + \epsilon_2, X \in R_1\} \\ &\vdots \\ \min_{X \in R_{p-1}} f_p(X) &= f_p^*(X_p), R_{p-1} = \{X | f_{p-1}(X) < f_{p-1}^* + \epsilon_{p-1}, X \in R_{p-2}\} \end{aligned}$$

则 X_p 为多目标问题最优解。

4.2.3 交互规划法

评价函数法是有局限性的。其一,这种方法给出的评价函数一般较粗糙,难以精确地描述它要预先给出的表达式;其二,在一些复杂问题中,客观存在的评价函数并不能简单写出;其三,评价函数法需要一开始就给出全部评价信息,而实际上这是十分困难的。很多情况下,评价信息只能在实际计算过程中逐步给出。鉴于此,就有必要改进这种求解方法,改进方向一是预先不给出评价函数的表达式;二是用迭代算法求解;三是求解过程中加入决策者的意见。

基于以上考虑,国际上从20世纪70年代开始提出了求解多目标问题的交互规划法。交互规划法是一种采用分析阶段和决策阶段反复交替进行的求解多目标最优化问题的方法,即是一种以分析者的求解和决策者的抉择相结合的人机对话式求解过程。其求解过程如图4.15所示。

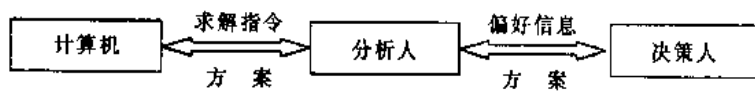


图 4.15

交互式方法的优点是,不需事先知道全部建模信息,而是可以在求解过程中逐步完善,决策者参与求解过程,根据自己的偏好向分析者提供信息,分析者根据信息不断向决策者提供方案,所得的解是使决策者满意的解。

1. 线性加权和型交互法

以两目标为例给出该方法的步骤:

①取权系数 $\lambda'_1 > 0, \lambda'_2 > 0, \lambda'_1 + \lambda'_2 = 1$, 求解

$$(P_1) \min_{X \in R} [\lambda'_1 f_1(X) + \lambda'_2 f_2(X)]$$

设其最优解为 X^1 , 相应目标值 $f_1^1 = f_1(X^1), f_2^1 = f_2(X^1)$;

②若决策者认为 f_2^1 太大(对第二目标值不满意),则增大 f_2 的权系数,不妨令 $\lambda_1 = \lambda'_1 - \Delta, \lambda_2 = \lambda'_2 + \Delta$, 其中 $0 < \Delta < \lambda'_1$, 同理,若决策者认为 f_1^1 不满意,则可类似地增大 f_1 的权系数;

③求解(计算机)

$$(P_2) \min_{X \in R} [(\lambda'_1 - \Delta) f_1(X) + (\lambda'_2 + \Delta) f_2(X)]$$

得最优解 X^2 , 相应目标值 $f_1^2 = f_1(X^2), f_2^2 = f_2(X^2)$, 可以证明,此时必有 $f_1^2 \geq f_1^1, f_2^2 \leq f_2^1$, 若决策者满意,则停止,否则转②。

2. 理想点型交互法(STEM 逐步法)

STEM 逐步法是一种迭代法。在求解过程中,每进行一步,分析者把计算结果告诉决策者,决策者对计算结果做出评价。若认为已满意了,则迭代停止,否则分析者再根据决策者的意见进行修改和再计算,如此,直到求得决策者认为满意的解为止,故称此法为逐步进行法或对话式方法。

逐步法要求多目标规划问题的目标函数 $f_i(X) (i=1, 2, \dots, p)$ 为线性函数,不妨设 $f_i(X)$

$= \sum_{j=1}^n c_{ij} x_j$, 则逐步法的求解步骤如下。

①(初始步)取初始约束集 $R^0 = R, k=0$ 。

②(分析步)首先通过求单目标规划问题,确定 R^k 上之理想点

$$F^* = [f_1^* \quad f_2^* \quad \cdots \quad f_p^*]$$

式中 $f_i^* = \min_{X \in R} f_i(X) \quad (i=1,2,\cdots,p)$

然后使用加权理想点法确定权系数 λ_i 。并求解

$$(P) \quad \min_{X \in R} \max_{1 \leq i \leq p} \lambda_i |f_i(X) - f_i^*|$$

或与之等价的辅助问题 (P)
$$\begin{cases} \min w \\ X \in R^k \\ \lambda_i [f_i(X) - f_i^*] \leq w \\ w \geq 0 \end{cases}$$

如令 $f_i^0 = \max_{1 \leq j \leq p} f_i(X^j)$, $X^j (j=1,2,\cdots,p)$ 为单目标最优点 (即把各单目标最优点一一代入 $f_i(X)$ 得一最差值 f_i^0), 则利用线性目标函数的信息, 权系数 λ_i 由下式给出:

$$\alpha_i = \begin{cases} \frac{f_i^0 - f_i^*}{f_i^0} \frac{1}{\sqrt{\sum_{j=1}^n c_{ij}^2}}, & \text{当 } f_i^0 > 0 \\ \frac{f_i^* - f_i^0}{f_i^*} \frac{1}{\sqrt{\sum_{j=1}^n c_{ij}^2}}, & \text{当 } f_i^0 \leq 0 \end{cases} \quad (i=1,2,\cdots,p)$$

经归一化后, 得权系数 $\lambda_i = \alpha_i / \sum_{i=1}^p \alpha_i$ 。其中, 上式的前一部分是目标 f_i 的最差(大)值和最优值的相对差, 它反映了对在约束集中变化敏感的目标赋以较大的权, 变化不敏感的目标赋以较小的权; 该式的后一部分是利用目标的“平均”系数使之标准化。

求解(P)得 k 步最优解 $[X^* \quad w^*]$, 则由理想点法知, 当 $\lambda_i > 0$ 时, X^* 为多目标极小化问题的弱有效解。

③(决策步)若记 $F^k = [f_1(X^k) \quad \cdots \quad f_p(X^k)]$, 问决策者对 F^k 满意否? 若全部满意, 则 $X^* = X^k$, 迭代停止。否则, 若有不满意的目标, 则对满意目标放宽(对极小型, 增大目标值), 放宽容量由决策者给出。不妨设决策者对 $f_{s_k}(X^k)$ 已满意, 其余不满意, 决策者给出的宽容度为 Δ_{s_k} , 即认为 $f_{s_k}(X^*) + \Delta_{s_k}$ 仍属满意, 则构成新的约束集合

$$R^{k+1} = \left\{ X \in R^k \left| \begin{array}{ll} f_{s_k}(X) \leq f_{s_k}(X^k) + \Delta_{s_k} & s_k \text{ 为满意目标下标} \\ f_j(X) \leq f_j(X^k) & j \text{ 为不满意目标下标} \end{array} \right. \right\}$$

置 $k+1=k$ 转②。

逐步法每次迭代调整的是约束集合 R^k , 决策者只要回答满意或不满意, 最终的迭代次数不超过 p 次, 因为方法中要求每一轮迭代至少使一个不满意的目标改善成满意。对于极大化问题的求解步骤, 读者完全可以仿照给出。

例 4.9 用 STEM 逐步法求解下述多目标极小化问题:

$$\min f_1(X) = -x_1$$

$$\min f_2(X) = -0.4x_1 - 0.3x_2$$

$$\text{s. t. } \begin{cases} x_1 + x_2 \leq 400 \\ 2x_1 + x_2 \leq 500 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

解 记 $\mathbf{X} = [x_1 \ x_2]^T$, 按 STEM 逐步法的求解步骤先进行分析阶段:

第 1 步, 分别对两个单目标进行极小化, 有

$$\begin{cases} \min f_1(\mathbf{X}) = -x_1 \\ \mathbf{X} \in R \end{cases} \text{ 得极小解 } \mathbf{X}^{(1)} = [250 \ 0]^T$$

$$\begin{cases} \min f_2(\mathbf{X}) = -0.4x_1 - 0.3x_2 \\ \mathbf{X} \in R \end{cases} \text{ 得极小解 } \mathbf{X}^{(2)} = [100 \ 300]^T$$

因为 $\mathbf{X}^{(1)} \neq \mathbf{X}^{(2)}$, 进行第 2 步, 计算各目标值得

$$f_1^* = f_1(\mathbf{X}^{(1)}) = 250 \quad f_{12} = f_1(\mathbf{X}^{(2)}) = -100 \quad f_{21} = f_2(\mathbf{X}^{(1)}) = -100$$

$$f_2^* = f_2(\mathbf{X}^{(2)}) = -130 \quad \text{则 } f_1^0 = \max\{f_1^*, f_{12}\} = -100 \quad f_2^0 = \max\{f_{21}, f_2^*\} = -100$$

由权系数求解公式, 有

$$\alpha_1 = \frac{-250 - (-100)}{-250} \left(\frac{1}{\sqrt{1^2 + 0^2}} \right) = 0.6$$

$$\alpha_2 = \frac{-130 - (-100)}{-130} \left(\frac{1}{\sqrt{0.4^2 + 0.3^2}} \right) = 0.4615$$

因而得权系数

$$\lambda_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2} = 0.5652, \lambda_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2} = 0.4348$$

求解辅助问题

$$\begin{cases} \min w \\ \mathbf{X} \in R \\ 0.5652(-x_1 + 250) \leq w \\ 0.4348(-0.4x_1 - 0.3x_2 + 130) \leq w \\ w \geq 0 \end{cases}$$

可得最优解

$$\bar{\mathbf{X}}^{(1)} = [230 \ 40]^T$$

相应的目标函数值

$$f_1(\bar{\mathbf{X}}^{(1)}) = -230 \quad f_2(\bar{\mathbf{X}}^{(1)}) = -104$$

现在进行第 3 步, 转入决策阶段。请决策者比较上述参考目标值和理想目标值 $f_1^* = -250$, $f_2^* = -130$ 。今设决策者认为对目标 f_1 已满意, 对目标 f_2 不满意。决策者为改善目标 f_2 , 考虑给出满意目标 f_1 的最大宽容量 $\Delta_1 = 30$ 。由此, 构造新约束集

$$R^1 = \left\{ \mathbf{X} \in R \mid \begin{cases} -x_1 \leq -230 + 30 \\ -0.4x_1 - 0.3x_2 \leq -104 \end{cases} \right\}$$

令 $\alpha_1 = 0$, 则 $\alpha_2 = 0.4615$, 故得新的权系数

$$\lambda_1 = 0 \quad \lambda_2 = 1$$

再求解下述辅助问题:

$$\begin{cases} \min w \\ \mathbf{X} \in R^1 \\ -0.4x_1 - 0.3x_2 + 130 \leq w \\ w \geq 0 \end{cases}$$

得最优解

$$\bar{\mathbf{X}}^{(2)} = [200 \quad 100]^T$$

相应的目标值

$$f_1(\bar{\mathbf{X}}^{(2)}) = -200 \quad f_2(\bar{\mathbf{X}}^{(2)}) = -110$$

否则,问题无满意解。

转入决策阶段,再请决策者比较上述参考目标值和理想目标值,设决策者认为对两个目标均已满意,则求解结束,得满意解

$$\bar{\mathbf{X}} = [200 \quad 100]^T$$

和相应的目标值

$$\bar{\mathbf{F}} = [-200 \quad -110]^T$$

利用交互规划法求解多目标规划问题只需决策者提供有关问题的局部偏爱信息,而不需预先知道偏爱的全部信息。同其他方法相比,交互规划有更少的限制性假设,决策者参与问题的解决过程,因而所得解是决策者满意的。不足之处在于,所得解过分依赖于决策者提供的局部偏爱信息的准确性,而且在不多的交互次数中,一般难以得到令人满意的解。

习 题

4.1 某厂生产 A、B、C 三种产品,有关数据如表 4.4。

表 4.4

产品	工厂生产能力(t/h)	单位利润(万元/t)	下月市场需求(t)
A	3	5	240
B	1	7	250
C	4	3	420

又知工厂每月正常开工能力为 208 h。市场需要尽可能多的 A 产品,问应如何安排下月的生产计划,可在避免开工不足的条件下,使工人加班时间尽量地少;工厂获得最大利润;满足市场对 A 产品的尽可能多的需求?试列出本问题的数学模型。

4.2 某社会团体筹建技术学院为企事业单位培训干部,招收专修班和研究班两类学员,学院现有专业课教师 36 人,基础课教师 42 人,专修班每 100 名学员需专业课教师 2 人,基础课教师 3 人;研究班每 100 名学员需专业课教师 6 人,基础课教师 4 人,学员由原单位缴纳培训费,专修班每人 4 千元,研究班每人 5 千元。根据设备条件,研究班人数不能超过专修班人数的一半。

现在学院需制定规划,即两类学员各录取若干人,设定的目标依次是:①培训费收入设定 5 000 千元,但超过更好;②专修班人数设定 1 000 人,越接近越好;③研究班人数设定 400 人,越接近越好。试列出该问题的目标规划模型。

4.3 在图 4.16 中标出有效解集合 R_{∞}^* ,弱有效解集合 R_{wp}^* 以及单目标的最优解集合 $R_1^* \cdot R_2^* \cdot R_3^*$ 。

4.4 用线性加权和法求解

$$(\text{VMP}) V = \min_{\mathbf{X} \in R} (f_1(\mathbf{X}) \cdot f_2(\mathbf{X}))$$

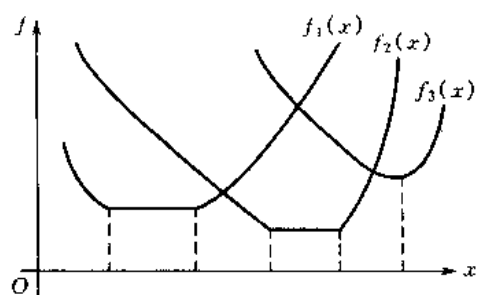


图 4.16

式中

$$\mathbf{X} = [x_1 \quad x_2]^T$$

$$f_1(\mathbf{X}) = -x_1 - 8x_2$$

$$f_2(\mathbf{X}) = -6x_1 - x_2$$

$$\text{s. t. } \begin{cases} 3x_1 + 8x_2 \leq 12 \\ x_1 + x_2 \leq 2 \\ 2x_1 \leq 3 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

$$\text{取 } \lambda_1 = \lambda_2 = \frac{1}{2}$$

4.5 在 4.1 题中,若要求总利润尽量接近 $f_1^0 = 3\,000$ 万元为第一优先,总工时尽量接近 $f_2^0 = 208$ h 和 A 产品产量尽量接近 $f_3^0 = 250$ t 为第二优先;同时决策者还建议总利润超过与不足 f_1^0 的权系数分别为 0.2 和 0.8;总工时必须尽可能达到或超过其目标值 f_2^0 ;A 产品产量超过与不足 f_3^0 的权系数分别为 0.4 和 0.6。试制订该厂下月的生产计划。

III 图与网络技术

第5章 图与网络分析

在现代管理中,经常碰到各式各样的图与网络,诸如交通道路图、生产流程图、煤气管道图等。这些图与网络,不论名称、内容如何,都具有两个基本要素:一是被研究的对象,如车站等,通常用点来表示;二是所研究对象之间的某种特定关系,如两车站之间的距离,通常用点与点之间的连线表示。图与网络理论就是撇开各种各样的具体内容来讨论这种由点、线组成的抽象形式的图,从中研究其一般规律以及典型问题的定性、定量分析方法和算法。

5.1 基本概念

图论中图是由点及点与点之间的连线构成,它可以反映一些对象之间的关系。

例如,甲、乙、丙、丁、戊五个足球队,它们之间比赛的情况,可以用图表示出来。已知甲队和其他各队都赛过一次,乙队和甲、丙队赛过,丙队和甲、乙、丁队赛过,丁队和甲、丙、戊队赛过,戊队和甲、丁队赛过。为了反映这个情况,可以用点 v_1, v_2, v_3, v_4, v_5 分别代表这五个队,某两个队之间比赛过,就在这两个队所相应的点之间联一条线,这条线不过其他的点,称为边,图论中的边通常用 e_i 表示,如图 5.1 所示。

从上面的例子可以看出,图可以很好地描述、刻画对象之间的特定关系。在一般情况下,图中点的相对位置如何,点与点之间连线的长短曲直,对于反映对象之间的关系并不是重要的。可见,图论中的图与几何图、工程图是不同的。

假如把上面的例子中“相互比赛”的关系改成“球队间的胜负关系”,那么,只用两点的连线就很难刻画他们之间的关系了。为了反映这一类关系,可以用一条带箭头的连线表示。如球队 v_1 胜了球队 v_2 ,可以从 v_1 引一条带箭头的连线到 v_2 。图 5.2 反映了五个球队比赛的胜负情况。可见, v_1 三胜一负, v_4 三场全负等等。类似胜负这种非对称性的关系,在生产和生活中是常见的,如城市交通道路中的“单行线”、“师生关系”等。

综上所述,一个图是由一些点及一些点之间的连线(带箭头或不带箭头)所组成。记为 G

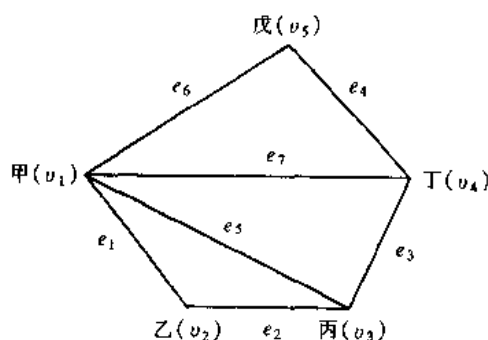


图 5.1

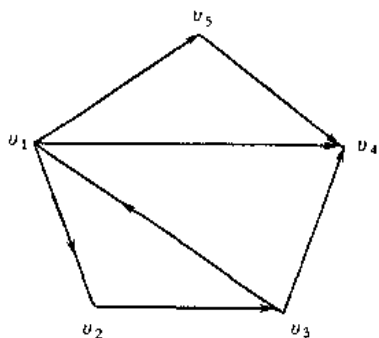


图 5.2

$= (V, E)$ 或 $D = (V, A)$ 。其中, $V = \{v_1, v_2, \dots, v_n\}$ 表示点的集合, $E = \{e_1, e_2, \dots, e_m\}$ 表示边的集合, $A = \{a_1, a_2, \dots, a_m\}$ 表示弧的集合。以下就结合上述例子介绍网络技术中一些常用的基本概念。

1. 边与弧、无向图与有向图

把两点之间不带箭头的连线称为边,带箭头的连线称为弧。若一个图 G 是由点及边构成,则称之为无向图,记为 $G = (V, E)$ 。一条联结点 $v_i, v_j \in V$ 的边记为 $e_i = [v_i, v_j]$ (或 $[v_j, v_i]$),称 e_i 是点 v_i, v_j 的关联边。如图 5.1 中,

$e_6 = [v_1, v_5]$ 。

若一个图 D 是由点及弧所构成,则称为有向图,记为 $D = (V, A)$ 。一条方向是从 v_i 指向 v_j 的弧记为 (v_i, v_j) 。如图 5.2 中, (v_1, v_2) 是一条从 v_1 出发到 v_2 的弧。

2. 链与路、圈与回路

在无向图中,若存在一个点边交错序列 $(v_{i_1}, e_{i_1}, v_{i_2}, e_{i_2}, \dots, v_{i_{k-1}}, e_{i_{k-1}}, v_{i_k})$, 满足 $e_{i_t} = [v_{i_t}, v_{i_{t+1}}]$ ($t=1, 2, \dots, k-1$), 则称之为一条联结 v_{i_1} 和 v_{i_k} 的链, 记为 $(v_{i_1}, v_{i_2}, \dots, v_{i_k})$ 。若 $v_{i_1} = v_{i_k}$, 则称之为圈。如在图 5.1 中, (v_1, v_2, v_3, v_4) 就是一条链, 而 $(v_1, v_2, v_3, v_4, v_5, v_1)$ 就是一个圈。

在有向图 D 中,若存在一个点弧交错序列 $(v_{i_1}, a_{i_1}, v_{i_2}, a_{i_2}, \dots, v_{i_{k-1}}, a_{i_{k-1}}, v_{i_k})$, 弧 a_{i_t} 的始点为 v_{i_t} , 终点为 $v_{i_{t+1}}$, 记为 $a_{i_t} = (v_{i_t}, v_{i_{t+1}})$, 则称这条点弧的交错序列为从 v_{i_1} 到 v_{i_k} 的一条路, 记为 $(v_{i_1}, v_{i_2}, \dots, v_{i_k})$ 。若路的第一点和最后一点相同, 则称之为回路。如在图 5.2 中, $(v_1, (v_1, v_2), v_2, (v_2, v_3), v_3, (v_3, v_4), v_4)$ 是从 v_1 到 v_4 的路, 而 $(v_1, (v_1, v_2), v_2, (v_2, v_3), v_3, (v_3, v_1), v_1)$ 则是一个回路。

链与路中的点互不相同, 则为初等链(路)。以后说到的链与路, 均指初等链(路), 类似可定义初等圈(回路)。

3. 连通图与支撑子图

图 G 中, 若任何两个点之间至少有一条链, 则称 G 是连通图, 否则称为不连通图, 如图 5.1 是一个连通图。

给定一个图 $G = (V, E)$, 若图 $G' = (V', E')$, 使 $V = V'$ 及 $E' \subseteq E$, 则称 G' 是 G 的一个支撑子图, 如图 5.3 就是图 5.1 的一个支撑子图。

4. 赋权图与网络

对于一个无向图 G 的每一条边, 或对于有向图 D 的每一条弧, 相应有一个权数 ω_{ij} (或 c_{ij}), 则称这样的图为赋权图。

网络一般是指一个弧上有某种所谓“流转物”流动的有向图。实际生活中, 这种具有流动物的有向图——网络是很多的, 常见的网络如表 5.1 所示。

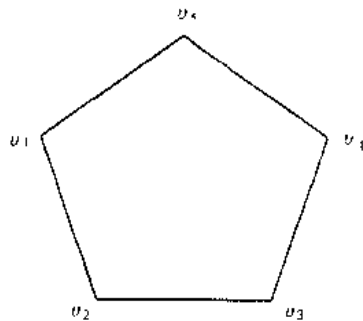


图 5.3

表 5.1

点(节点)	弧(分支)	流转物
交叉口	公路	车辆
机场	空中航线	飞机
泵站	管道	液体
电报局	电缆	电报
中继站	通信线路	信息

在实际中,每个弧上对流动的流转物都有流转能力限制。例如,每段公路上都有最大车流量的限制,每根管道上都有通过最大流量的限制,把这种流转能力的限制称为弧上的容量。另外,在实际问题中,通常有一点是发点,另有一点是收点,其余为中间点。这样就给出网络的如下定义:在有向图 $D=(V,A)$ 中, v_s 为起点, v_t 为终点,而对每一弧 $(v_i, v_j) \in A$ 赋以量 $c_{ij} > 0$ 称为弧的容量,则称这样的有向图为一个网络,记为 $D=(V,A,C)$ 。

5.2 最小支撑树问题

在各式各样的图中,有一类图是极其简单然而却是很有用的,这就是树。所谓树就是一个无圈的连通图。在树图中,只要任意再加上一条边,必然会出现圈,而树图中任何二个节点之间有一条且仅存一条惟一的链。也就是说,若去掉任一条边,那么连通图就被破坏,因此树图是边数最少的连通图。

设图 $T=(V,E')$ 是图 $G=(V,E)$ 的支撑子图。若图 $T=(V,E')$ 是一个树图,则称 T 是 G 的一个支撑树。

所谓最小支撑树问题就是在一个赋权的连通的无向图 G 中找出一个支撑树,并使得这个支撑树的所有边的权数之和为最小。

这里所说的“权”,是指与边有关的数量指标。根据实际问题的需要,可以赋予它不同的含义,如距离、时间、费用等。

5.2.1 求解最小支撑树问题的破圈法

这种算法的具体步骤为:

- ①在给定的赋权的连通图上任找一个圈;
- ②在所找的圈中去掉一条权数最大的边(若有两条或两条以上的边都是权数最大的边,则任意去掉其中一条);
- ③若所余下的图已不含圈,则计算结束,所余下的图即为最小支撑树,否则,返回①。

例 5.1 某大学准备将所属 7 个学院办公室的计算机联网。这个网络的可能联通的途径如图 5.4 所示。图中 v_1, \dots, v_7 表示 7 个学院办公室,图中的边为可能联网的途径,边上所赋权数为这条路线的长度,单位为百米。试设计一个网络联通 7 个学院办公室,并使总的线路长度为最短。

解 这个问题实际上是求图 5.4 的最小支撑树,具体步骤为:

- ①在图 G 中找到一个圈 (v_1, v_6, v_7, v_1) , 去掉其中权最大的边 (v_1, v_6) , 得图 G_1 ;
- ②在图 G_1 中找到一个圈 $(v_3, v_4, v_5, v_7, v_3)$, 去掉其中权最大的边 (v_4, v_5) , 得图 G_2 ;

- ③在图 G_2 中找到一个圈 $(v_2, v_3, v_5, v_7, v_2)$, 去掉其中权最大的边 (v_5, v_7) , 得图 G_3 ;
- ④在图 G_3 中找到一个圈 $(v_3, v_5, v_6, v_7, v_3)$, 去掉其中权最大的边 (v_5, v_6) , 得图 G_4 ;
- ⑤在图 G_4 中找到一个圈 (v_2, v_3, v_7, v_2) , 去掉其中权最大的边 (v_3, v_7) , 得图 G_5 ;
- ⑥在图 G_5 中已找不到任何一个圈了, 则 G_5 即为图 G 的最小支撑树, 如图 5.5 所示。

这个最小支撑树的所有边的总权数为 $3 + 3 + 3 + 1 + 2 + 7 = 19$, 即该校若按图 5.5 的设计, 可使此网络的总的线路长度最短, 共 1900 米。

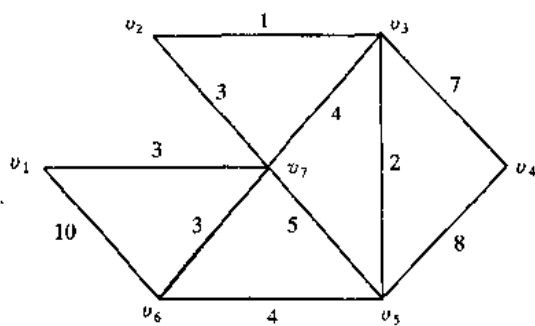


图 5.4

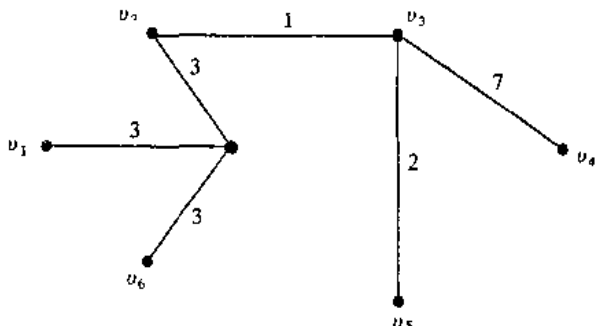


图 5.5

5.2.2 求解最小支撑树问题的避圈法

这种算法的具体步骤为:

- ①在赋权图中, 任选一点 v_i , 找出所有与 v_i 相关联的权最小的边 (v_i, v_j) , 得点 v_j ;
- ②把所有顶点分为互补的二部分 X, \bar{X} , 其中 X 表示与已选的边相关联的点集, \bar{X} 是不与已选的边相关联的点集;

- ③考虑这样的边 (v_i, v_j) , 其中 $v_i \in X, v_j \in \bar{X}$, 挑选其中权最小的边;

- ④重复第③步, 直至全部顶点皆属于 $X (\bar{X} = \emptyset)$ 。

例 5.2 用避圈法求下述赋权图(图 5.6)的最小支撑树。

解 ①任选一点 v_1 , 挑与之相关联的权最小的边 (v_1, v_2) 。

- ②所有顶点可分为互补的二部分: $X = \{v_1, v_2\}, \bar{X} = \{v_3, v_4, v_5, v_6\}$, 从边 $(v_2, v_4), (v_2, v_5), (v_2, v_3), (v_1, v_3)$ 中挑其中权最小的边 (v_2, v_3) 。

- ③顶点集分为互补的二部分: $X = \{v_1, v_2, v_3\}, \bar{X} = \{v_4, v_5, v_6\}$ 。从边 $(v_3, v_5), (v_2, v_5), (v_2, v_4)$ 中, 挑选其中权最小的边 (v_2, v_4) 。

- ④顶点集分为互补的二部分: $X = \{v_1, v_2, v_3, v_4\}, \bar{X} = \{v_5, v_6\}$ 。从边 $(v_4, v_6), (v_2, v_5), (v_4, v_5), (v_3, v_5)$ 中, 挑选其中权最小的边 (v_4, v_5) 。

- ⑤顶点集分为互补的二部分: $X = \{v_1, v_2, v_3, v_4, v_5\}, \bar{X} = \{v_6\}$ 。从边 $(v_4, v_6), (v_5, v_6)$ 中, 挑选其中权最小的边 (v_5, v_6) , 由于此时 $\bar{X} = \emptyset$, 所以, 经上述步骤所挑选的边构成的图, 即为图 5.6 的最小支撑树, 如图 5.7 所示。

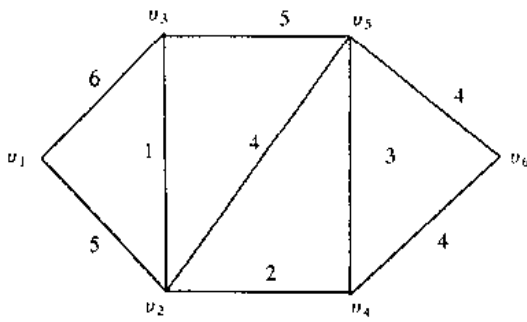


图 5.6

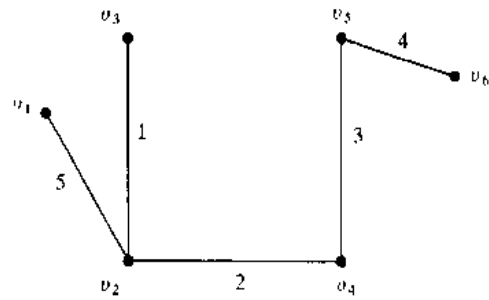


图 5.7

5.3 最短路问题

最短路问题是在一个网络(赋权有向图)中寻找从起点到某个节点之间一条最短的路线。图 5.8 表示城市 v_1 至城市 v_6 之间的公路系统的网络。弧上的数字表示公路长度,现欲求出 v_1 至 v_6 距离最短的路径。

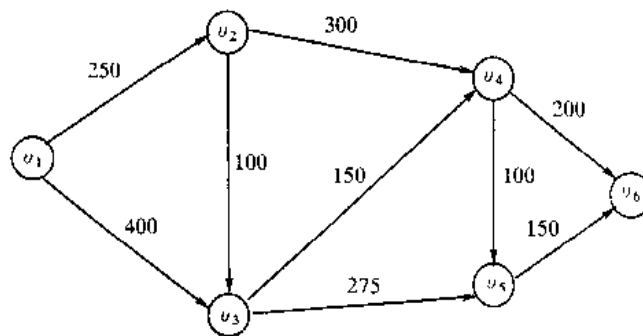


图 5.8

求解上述最短路问题,一种简单想法是找出 v_1 至 v_6 所有可能的通路,比较路的长度,确定路长最短的一条。但实际问题网络比较复杂,一一枚举所有通路的工作量往往非常大,这样枚举方法是行不通的,只有利用一些特殊的方式求解。

5.3.1 最短路问题的 Dijkstra 解法

Dijkstra 方法是求解最短路问题的有效方法之一。其基本思想是:从起点 v_1 开始逐步计算从 v_1 到网络各中间点 v_i 的最短路,逐步外推直至算出 v_1 至终点 v_n 的最短路。算法过程可直接通过在网络图上逐步标号完成,如果已求出 v_1 至 v_i 的最短路,即可给点 v_i 标上号 $[\alpha_i, \beta_i]$ 。第一个标号 α_i 表示 v_1 至 v_i 点最短路的的路长,第二个标号 β_i 表示 v_1 至 v_i 的最短路中 v_i 之前接点为 v_{β_i} ,若给终点 v_n 标上号 $[\alpha_n, \beta_n]$,则表示已求出 v_1 至 v_n 的最短路,其最短路长为 α_n ,最短路径可根据标号 β_i 反向追踪而得。

算法基本步骤如下。

①首先对起点 v_1 标号 $[\alpha_1, \beta_1]$,即计算 v_1 至 v_1 的最短路,最短路长为零,标号 $\alpha_1=0, \beta_1=1$ 。

②将网络中所有顶点分成已标号和未标号两类。 $v_i \in X$ 表示 v_i 是已标号点, 即已求出 $v_1 \rightarrow v_i$ 的最短路。 $v_j \in \bar{X}$ 表示 v_j 是尚未标号点, 即尚未求出 $v_1 \rightarrow v_j$ 的最短路。考虑网络中所有这样的弧

$$(v_i, v_j) \in (X, \bar{X})$$

即由已标号点 v_i 出发流向未标号点 v_j 的弧, 若这样的弧不存在, 即集合 (X, \bar{X}) 是空集时, 表示所有点都已标号, 即终点 v_t 也标上号, 算法完成, v_1 至 v_t 的最短路径可由 v_t 的第二标号起反向追踪而得, 即 v_t 前一点是 v_{β_t} , 再由 v_{β_t} 的第二个标号确定 v_{β_t} 之前一点, 一直反推至 v_1 , 就得 v_1 至 v_t 的一条最短路, 其路长为 α_t 。

③若 (X, \bar{X}) 非空, 计算

$$\min_{(v_i, v_j) \in (X, \bar{X})} (\alpha_i + w_{ij}) = \alpha_{i_k} + w_{i_k j_k}$$

其中 w_{ij} 为弧 (v_i, v_j) 之长度, 得一未标号顶点 v_{j_k} 。

④对顶点 v_{j_k} 标号 $[\alpha_{j_k}, \beta_{j_k}]$, 其中

$$\alpha_{j_k} = \alpha_{i_k} + w_{i_k j_k}, \beta_{j_k} = i_k$$

表示 v_1 至 v_{j_k} 最短路已求出, v_{j_k} 点由未标号点变成已标号点, 重复②。

现以图 5.9 所示最短路问题为例说明上述算法的具体步骤。

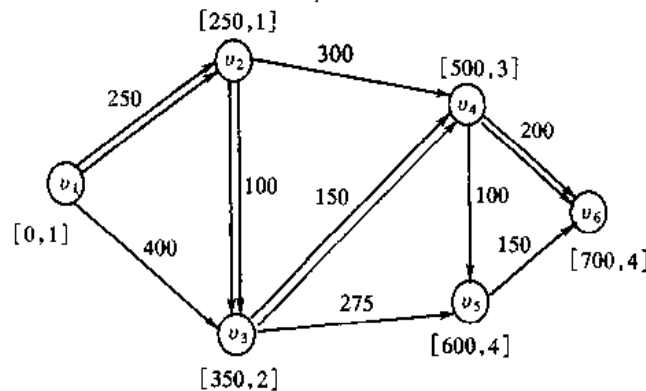


图 5.9

①给 v_1 标号 $[0, 1]$ 。

② $X = \{v_1\}$, $\bar{X} = \{v_2, v_3, v_4, v_5, v_6\}$ 。检查弧 (v_1, v_2) 、 (v_1, v_3) , 计算

$$\min\{0 + 250, 0 + 400\} = 250$$

则给 v_2 标号 $[250, 1]$ 。

③ $X = \{v_1, v_2\}$, $\bar{X} = \{v_3, v_4, v_5, v_6\}$, 检查弧 (v_1, v_3) 、 (v_2, v_3) 、 (v_2, v_4) , 计算

$$\min\{0 + 400, 250 + 100, 250 + 300\} = 350$$

则给 v_3 标号 $[350, 2]$ 。

④ $X = \{v_1, v_2, v_3\}$, $\bar{X} = \{v_4, v_5, v_6\}$, 检查弧 (v_2, v_4) 、 (v_3, v_4) 、 (v_3, v_5) , 计算

$$\min\{250 + 300, 350 + 150, 350 + 275\} = 500$$

则给 v_4 标号 $[500, 3]$ 。

⑤ $X = \{v_1, v_2, v_3, v_4\}$, $\bar{X} = \{v_5, v_6\}$, 检查弧 (v_3, v_5) 、 (v_4, v_5) 、 (v_4, v_6) , 计算

$$\min\{350+275, 500+100, 500+200\} = 600$$

则给 v_5 标号 $[600, 4]$ 。

⑥ $X = \{v_1, v_2, v_3, v_4, v_5\}$, $\bar{X} = \{v_6\}$, 检查弧 (v_4, v_6) , (v_5, v_6) , 计算

$$\min\{500+200, 600+150\} = 700$$

则给 v_6 标号 $[700, 4]$, 算法完成。

由 v_6 之第二个标号起反向追踪得

$$v_6 \leftarrow v_4 \leftarrow v_3 \leftarrow v_2 \leftarrow v_1$$

即得由 v_1 至 v_6 的最短路为

$$v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_4 \rightarrow v_6$$

最短路长 $\alpha_6 = 700$ 。

5.3.2 最短路问题模型的应用

最短路问题模型不仅能解决寻找最短距离问题,对一些其他类型的应用问题,如设备更新、投资决策、计划安排等也能求解。在这些问题中,网络中的弧长 w_{ij} 不再表示距离,而是表示费用、时间等经济量。

例 5.3(设备更新问题) 某厂使用一种设备,每年年初设备科需要对该设备的更新与合作出决策。若购置新设备,就要支付购置费用;若继续使用旧设备,则需支付维修费,设备使用的年数越长,每年所需的维修费就越多。现若第一年年初购置了一台新设备,问在 5 年内如何制定设备更新计划,以便使新设备购置费用和旧设备维修费的总费用最小? 已知设备在 5 年内各年初的价格(表 5.2)及设备使用不同年数的维修费(表 5.3)。

表 5.2

第 i 年	1	2	3	4	5
价格 a_i	11	11	12	12	13

表 5.3

使用寿命	0~1	1~2	2~3	3~4	4~5
费用 b_j	b_1	b_2	b_3	b_4	b_5
	5	6	8	11	18

解 首先建立一个网络模型(图 5.10)描述此更新问题。其中

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$$

$$E = \{(v_i, v_j), i=1, 2, \dots, 5; j=2, \dots, 6, i < j\}$$

节点 v_1, v_2, \dots, v_5 相当于第 1、2、 \dots 、5 年的年初, v_6 相当于第 5 年年末。

网络图中的弧 (v_i, v_j) 相当于第 i 年年初购买一台设备一直用到第 $j-1$ 年年末。

网络图中每一条从 v_1 至 v_6 的路表示一种可能的设备更新方案,例如路 $v_1 v_3 v_5 v_6$ 相当于第 1、3、5 年初购买设备这一方案。

网络弧上的数可按如下法则给出:

$$\begin{aligned} (v_i, v_j) \text{ 上数 } w_{ij} &= \text{第 } i \text{ 年初设备购买费} + (j-i) \text{ 年里的维修费} \\ &= a_i + (b_1 + b_2 + \dots + b_{j-i}) \end{aligned}$$

例如,弧 (v_2, v_5) 表示第2年初购买一台新设备,使用至第4年底,相应购买费为11,维修费为 $5+6+8=19$,弧 (v_2, v_5) 上的数 $w_{2,5}=30$ 。

这样,制定一个最优设备更新方案的问题就等价于寻求图5.10中 v_1 至 v_6 的最短路问题,使用Dijkstra算法得上述问题的最短路为 $v_1 v_4 v_6$,即第1、4年初购买设备,五年最佳总费用为53。

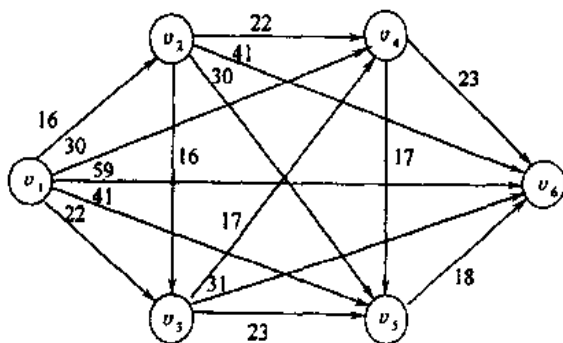


图 5.10

例 5.4(生产安排问题) 某公司欲尽快生产出一新产品去占领市场。研制这新产品需经过设计、试制、生产及调拨三个阶段,而每阶段可以以正常和应急二种水准进行。每阶段在这二种水准上所需时间由表5.4给出。现有8万元经费可供使用,每阶段中采用不同水准进行的费用如表5.5所示。试确定在预算经费的限制下,每阶段按什么水准进行才能使新产品投放市场的时间最短?

表 5.4

(天)

	设计	试制	生产与调拨
正常	4	3	5
应急	2	2	3

表 5.5

(万元)

	设计	调制	生产与调拨
正常	2	2	3
应急	3	3	4

解 对于所述问题可用图5.11所示的网络描述,每条弧 (v_i, v_j) 表示某阶段以某种水准进行工作。例如, (v_1, v_2) 表示以正常水准进行设计阶段工作, (v_4, v_9) 表示以应急水准进行生产与调拨阶段工作。

节点 v_1 表示整个工作开始,节点 v_{12} 表示整个工作完成。每个节点 v_i 上括号中数字 (S_i) 表示从 v_i 开始尚有经费 S_i 万元可使用。

弧 (v_i, v_j) 上所标 $w_{ij}(b_{ij})$ 表示 w_{ij} 是完成 (v_i, v_j) 的时间, b_{ij} 表示完成 (v_i, v_j) 所需的费用,满足 $S_i - b_{ij} = S_j$ 。而弧 (v_8, v_{12}) 、 (v_9, v_{12}) 、 (v_{10}, v_{12}) 、 (v_{11}, v_{12}) 并不表示实际工作。它们表示添加的虚工作,只是为了使网络有一个共同的终点,所以这些弧上 $w_{ij}=0, b_{ij}=0$ 。

上述问题每一个可行的计划方案都表示为网络中 v_1 至 v_{12} 的一条通路,而网络中每一通路也表示一个可行的计划方案(不超过预算费)。例如, $v_1 \rightarrow v_2 \rightarrow v_5 \rightarrow v_{10} \rightarrow v_{12}$ 表示正常水准设

计,应急水准试制,正常水准生产与调拨。这样,上述寻求最佳计划安排问题就转化为网络中求 v_1 至 v_{12} 的最短路问题。使用标号算法得出最短路为

$$v_1 \rightarrow v_3 \rightarrow v_6 \rightarrow v_{11} \rightarrow v_{12}$$

$$v_1 \rightarrow v_2 \rightarrow v_4 \rightarrow v_9 \rightarrow v_{12}$$

最短路长为 10。即最优计划安排有二个:

- ① 应急水准设计—正常水准试制—正常水准生产;
- ② 正常水准设计—正常水准试制—应急水准生产。

两种方案均需用时间 10 天。

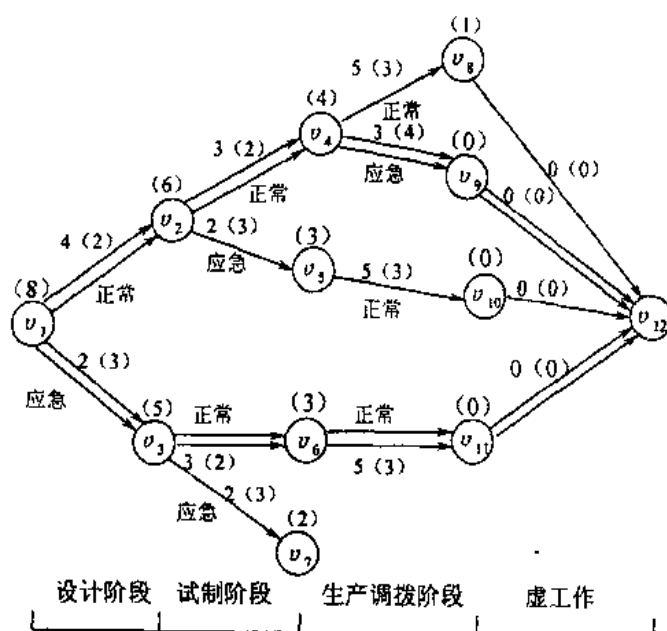


图 5.11

5.4 最大流问题

在研究网络问题时,经常遇到在网络系统中运送“流转物”的问题。交通系统网络中,每条弧表示一段通路,每一个节点则表示交叉道路,网络中运送的流转物为车辆。通信系统网络中,节点表示中转站,弧表示中继站之间的通信线路,网络中运送的流转物是信息。无论是公路系统还是通信系统,每段弧上允许通过流转物的能力是有限的。这种限制称为弧上的容量,弧 (v_i, v_j) 上容量用 c_{ij} 表示。现在讨论只有一个发点和一个收点的网络系统,假设要把起点的一批流转物运送到终点去,在每一弧上通过流转物的总量不能超过这条弧上的容量,问题是应该怎样安排运送,才能使从起点运送至终点的总量达到最大? 这样的问题就称为网络上最大流问题。

图 5.12 表示一个典型的最大流问题。图中网络表示城市 v_1 和城市 v_6 之间的公路系统,其中 v_2, v_3, v_4, v_5 是公路系统内的 4 个城市,弧上的数字表示容量,问题是如何确定每条公路 (v_i, v_j) 上通过的运输量,使 v_1 至 v_6 的运输量最大。

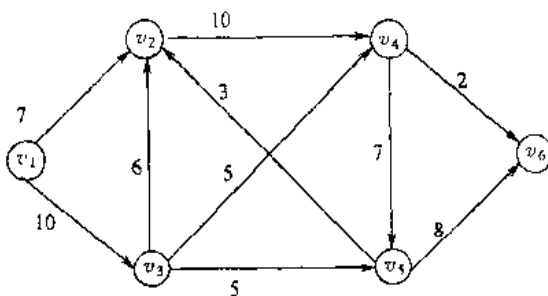


图 5.12

5.4.1 网络流的基本概念与基本定理

考虑网络最大流问题时,只关心通过网络的流转物的总量及其通过方式,即每段弧上应分配通过多少量,至于流通物本身是什么是无关紧要的,这样就产生网络中抽象的“流”的概念。

以下先介绍网络流的有关概念。

1. 流

这是指定义在弧集 $A = \{(v_i, v_j)\}$ 上的一个函数 $f = f(v_i, v_j)$, 通常记为 $|f_{ij}|$, 每个 f_{ij} 称为弧 (v_i, v_j) 上的流量。

2. 可行流

$|f_{ij}|$ 为网络上一个流, 它应满足容量条件和平衡条件。

1) 容量条件 对每一弧 $(v_i, v_j) \in A$, $0 \leq f_{ij} \leq c_{ij}$ 。

2) 平衡条件 经过中间点的流入量等于流出量, 对中间点 v_i 需满足:

$$\sum_{(v_i, v_j) \in A} f_{ij} - \sum_{(v_j, v_i) \in A} f_{ji} = 0$$

若网络起点为 v_s , 终点为 v_t , 则一个可行流的总流量为

$$\begin{aligned} \sum_{(v_s, v_j) \in A} f_{sj} - \sum_{(v_j, v_s) \in A} f_{js} &= V(f) \\ \sum_{(v_i, v_t) \in A} f_{it} - \sum_{(v_j, v_t) \in A} f_{jt} &= -V(f) \end{aligned}$$

式中 $V(f)$ 称为这个可行流的流量, 即发点的净输出量(或收点的净输入量)。

对任何一个网络总存在可行流, 如令所有弧的流量 $f_{ij} = 0$, 就得到一个可行流(零流)。一个网络上所有可行流中流量最大的可行流称为最大流。所以, 网络的最大流问题即是寻找流量最大的可行流。

3. 饱和弧与零流弧

对网络上一个可行流 $|f_{ij}|$, 若某弧上 (v_i, v_j) 的流量 $f_{ij} = c_{ij}$ (容量), 则称该弧为饱和弧, 否则称其为非饱和弧。

对一个可行流 $|f_{ij}|$, 若某弧上分配流量 $f_{ij} = 0$, 称该弧为零流弧, 否则称其为非零流弧。

4. 可扩充链(或称增广链)

设 $|f_{ij}|$ 是网络上可行流, 若存在起点 v_s 与终点 v_t 之间一条链 μ , 规定其方向为 $v_s \rightarrow v_t$ 。链上相应弧按 μ 方向分成两类: 与 μ 方向一致的弧为前向弧, 相反的弧则为后向弧。若

链 μ 上弧的流量满足:① μ 上所有前向弧上流量 $f_{ij} < c_{ij}$;② μ 上所有后向弧上流量 $f_{ij} > 0$, 则称 μ 为关于流 $\{f_{ij}\}$ 的可扩充链。

图 5.13 表示图 5.12 所示网络的一条链 μ :

$$v_1 - v_2 - v_3 - v_4 - v_5 - v_6$$

其中前向弧为 $\{(v_1, v_2), (v_3, v_4), (v_4, v_5), (v_5, v_6)\}$, 后向弧为 $\{(v_3, v_2)\}$ 。

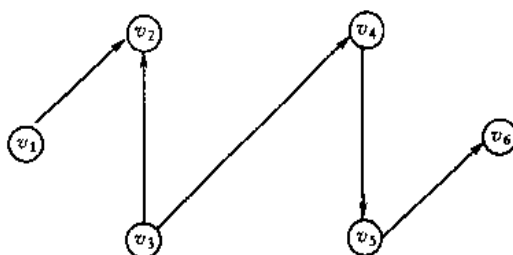


图 5.13

5. 截集与截量

对于网络 $D=(V, A, C)$, 若点集 V 被分割为两个非空集合 V_1 和 \bar{V}_1 , 使 $v_s \in V_1, v_t \in \bar{V}_1$, 则把弧集 (V_1, \bar{V}_1) (即始点在 V_1 , 终点在 \bar{V}_1 的所有弧构成的集合) 称为是 (分离 v_s 和 v_t) 的截集。从直观上讲, 截集是从 v_s 至 v_t 的必经之道。若把某一截集的弧从网络中去掉, 则从 v_s 至 v_t 便不存在路。一截集 (V_1, \bar{V}_1) 中所有弧的容量之和称为这个截集的容量, 简称为截量, 记为 $C(V_1, \bar{V}_1)$

$$C(V_1, \bar{V}_1) = \sum_{(v_i, v_j) \in (V_1, \bar{V}_1)} c_{ij}$$

显然, 任何一个可行流的流量 $V(f)$ 都不会超过任一截集的容量, 即 $V(f) \leq C(V_1, \bar{V}_1)$ 。因此, 对于一个可行流 f^* , 网络中有一个截集 (V_1^*, \bar{V}_1^*) , 使 $V(f^*) = C(V_1^*, \bar{V}_1^*)$, 则 f^* 必定是最大流, 而 (V_1^*, \bar{V}_1^*) 也必定是所有截集中容量最小的一个, 即最小截集。

易证, 可行流 f^* 是最大流, 当且仅当网络中不存在关于 f^* 的可扩充链。从而有下述的最大流量最小截集定理: 任一网络 D 中, 从 v_s 至 v_t 的最大流的流量等于分离 v_s, v_t 的最小截集的容量 (证法略)。

对于一个可行流 $\{f_{ij}\}$, 若能找到一条可扩充链, 则一定可以调整成一个流量更大的可行流 $\{\bar{f}_{ij}\}$, 首先对该链确定一调整量

$$\theta = \min\{\theta_1, \theta_2\}$$

式中, $\theta_1 = \min\{c_{ij} - f_{ij} \mid (v_i, v_j) \text{ 是链 } \mu \text{ 上前向弧}\}$; $\theta_2 = \min\{f_{ij} \mid (v_i, v_j) \text{ 是链 } \mu \text{ 上后向弧}\}$ 。

若链上无后向弧, 取 $\theta_2 = \infty$, 显然 $\theta > 0$ 。然后可确定新的可行流 \bar{f}_{ij} , 取

$$\bar{f}_{ij} = \begin{cases} f_{ij}, & (v_i, v_j) \notin \mu \\ f_{ij} + \theta, & (v_i, v_j) \text{ 是 } \mu \text{ 上前向弧} \\ f_{ij} - \theta, & (v_i, v_j) \text{ 是 } \mu \text{ 上后向弧} \end{cases}$$

容易验证, \bar{f}_{ij} 是网络上一个可行流, 若可行流 $\{f_{ij}\}$ 的流量为 $V(f)$, 则新可行流的流量 $V(\bar{f}) = V(f) + \theta$ 。进一步, 若 D 中仍有关于 \bar{f} 的可扩充链, 则可按上述方法, 继续改进 \bar{f} , 否则, 由上述结论可知, 此时已得到最大流, 同时也得到了一个最小截集。

6. 最大流问题的数学模型

由前面的讨论可知, 最大流问题就是在一个带收发点的网络中, 在不超过每条弧的容量的前提下求出从发点到收点的最大流量。这显然是一个线性规划问题, 数学模型的一般形式如下:

$$\begin{cases} \max V(f) \\ \sum_{(v_i, v_j) \in A} f_{ij} - \sum_{(v_j, v_i) \in A} f_{ji} = 0 \\ \sum_{(v_s, v_j) \in A} f_{sj} - \sum_{(v_j, v_s) \in A} f_{js} = V(f) \\ \sum_{(v_i, v_t) \in A} f_{it} - \sum_{(v_j, v_i) \in A} f_{ji} = -V(f) \\ 0 \leq f_{ij} \leq c_{ij} \\ i, j \neq s, t \end{cases}$$

例如,有下列一个石油运输管道图(图 5.14),某公司欲采用这个网络系统从 v_1 向销地 v_7 运送原油,弧的容量 c_{ij} 已给定(c_{ij} 的单位为万加仑/时,因管道直径的变化, c_{ij} 是不一样的)。问如何在每一弧上分配流量,方能使每小时运送尽量多的原油?

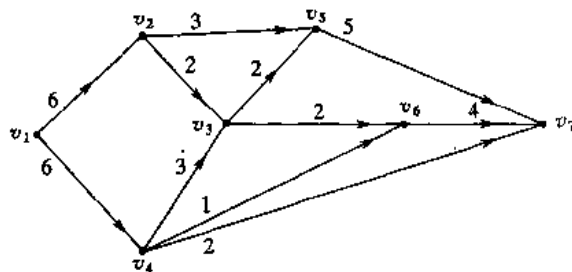


图 5.14

可以以此为例建立数学模型。设弧 (v_i, v_j) 上的流量为 f_{ij} , 网络上的总的流量为 F , 则有

目标函数 $\max F = f_{12} + f_{14}$

$$\text{约束条件: } \begin{cases} f_{12} = f_{23} + f_{25} \\ f_{14} = f_{43} + f_{46} + f_{47} \\ f_{23} + f_{43} = f_{35} + f_{36} \\ f_{25} + f_{35} = f_{57} \\ f_{36} + f_{46} = f_{67} \\ f_{57} + f_{67} + f_{47} = f_{12} + f_{14} \\ f_{ij} \leq c_{ij} \quad (i=1, 2, \dots, 6; j=2, \dots, 7) \\ f_{ij} \geq 0 \quad (i=1, 2, \dots, 6; j=2, \dots, 7) \end{cases}$$

在这个线性规划模型中,其约束条件的前 6 个方程表示了网络中的流量必须满足平衡条件,发点的总流出量必须等于收点的总流入量。其余的点称之为中间点。它的总流入量必须等于总流出量。后面几个约束条件表示对每一条弧 (v_i, v_j) 的流量 f_{ij} 要满足流量的可行条件,应小于等于弧 (v_i, v_j) 的容量 c_{ij} , 并大于等于零,即 $0 \leq f_{ij} \leq c_{ij}$ 。把满足平衡条件及流量可行条件的一组网络流 $\{f_{ij}\}$ 称之为可行流(即线性规划模型的可行解),可行流中一组流量最大(也即发点总流出量最大)的称之为最大流(即线性规划的最优解)。

把上述数据 c_{ij} 代入以上线性规划模型,计算后马上得到以下的结果:

$$f_{12} = 5 \quad f_{14} = 5 \quad f_{23} = 2 \quad f_{25} = 3 \quad f_{43} = 2 \quad f_{46} = 1$$

$$f_{47}=2 \quad f_{35}=2 \quad f_{36}=2 \quad f_{57}=5 \quad f_{67}=3$$

最优值(最大流量)=10。

对于网络最大流问题的求解,当然可以使用线性规划中的单纯形解法。但对于这一问题,实际中有简便算法,即标号法。

5.4.2 最大流问题的标号解法

寻找网络最大流的标号解法实际上分为两个过程,一是标号过程,即寻找可扩充链的过程;二是调整过程,即对当前的可行流进行改进的过程。其主要步骤是:

①确定一初始可行流 $\{f_{ij}\}$ 及其流量 $V(f)$;

②检验当前所确定可行流是否是网络中的最大流,若不是,需进一步调整;(正如前面所分析的那样,检验一个可行流是否为最大流,只要检查一下当前可行流是否还存在可扩充链,若存在,则说明当前可行流还不是最大流,否则是最大流。)

③将当前的可行流调整成一个流量更大的新可行流,再由②检验。

通常用观察法确定网络的一个初始可行流。对于较为复杂的网络,至少能把初始可行流取为零流。

通过在网络上标号的方法能系统地找出当前可行流的可扩充链。它的基本思想是:从起点 v_s 起,逐步寻找 v_s 至各点 v_i 间的可扩充链,若能找到 v_s 至 v_i 的一条可扩充链,则给点 v_i 标号 $[\theta_i, \alpha_i]$ 。第一个标号 θ_i 即为 v_s 至 v_i 这条可扩充链上的最大可调整量,第二个标号 α_i 则表示这条链上点 v_i 的前一点是 α_i 点。根据标号可反向追踪而写出这条链。在逐步扩大已标号的过程中,一旦终点 v_t 标上号,表示已找到一条由 v_s 至 v_t 的可扩充链。反之,如果标号过程进行至某一步中止了,而 v_t 尚未标号,则表明对当前的可行流,网络中不存在任何可扩充链。标号法的具体步骤如下。

①给发点 v_s 标号 $[\infty, v_s]$,含义为 v_s 至 v_s 的可扩充链已找到,前一点为 v_s ,这链上的可调整量为 ∞ 。选与 v_s 关联的从 v_s 流出的未饱和弧 (v_s, v_i) 或流入 v_s 的非零流弧 (v_i, v_s) ,给 v_i 标号 $[\theta_i, v_s]$ (对于流出弧)或 $[\theta_i, -v_s]$ (对于流入弧)。其中, $\theta_i = \begin{cases} c_{si} - f_{si} & \text{若}(v_s, v_i)\text{为流出未饱和弧;} \\ f_{si} & \text{若}(v_i, v_s)\text{为流入非零弧。} \end{cases}$

②把顶点集分为互补的二个点集 V_1, \bar{V}_1 ,其中 V_1 为已标号点集, \bar{V}_1 为未标号点集。

③考虑所有这样的弧 (v_i, v_j) 或 (v_j, v_i) ,其中 $v_i \in V_1, v_j \in \bar{V}_1$ 。若弧 (v_i, v_j) 为从 v_i 流出的未饱和弧,则给 v_j 标 $[\theta_j, v_i]$ 。其中

$$\theta_j = \min\{\theta_i, c_{ij} - f_{ij}\}$$

若弧 (v_j, v_i) 为流入 v_i 的非零流弧,则给 v_j 标号 $[\theta_j, -v_i]$,其中

$$\theta_j = \min\{\theta_i, f_{ji}\}$$

依此进行,得到的结局是:

(a) $\bar{V}_1 = \Phi$,说明网络中存在可扩充链 μ ,则由标号点反向追踪找出 μ ,转第④步;

(b) $\bar{V}_1 \neq \Phi$,标号已进行不下去,说明对于当前可行流,网络中已无可扩充链,得最大流 f^* ,同时得到最小截集 (V_1, \bar{V}_1) 。

④调整过程。取 $\theta = \min_{v_j \in \mu} \{\theta_j\}$, 令

$$\bar{f}_{ij} = \begin{cases} f_{ij} + \theta(v_i, v_j) & \text{是 } \mu \text{ 上前向弧} \\ f_{ij} - \theta(v_j, v_i) & \text{是 } \mu \text{ 上后向弧} \\ f_{ij} & \text{其他} \end{cases}$$

得新可行流 $\{\bar{f}_{ij}\}$ 。其流量 $V(\bar{f}) = V(f) + \theta$, 即比原可行流流量增加了 θ , 再转第①步。

例 5.5 有三个发电站(节点 v_1, v_2, v_3), 发电能力分别为 15、10 和 40 兆瓦, 经输电网可把电力送到 8 号地区(节点 v_8), 电网的能力如图 5.15。求三个发电站输到这地区(节点 v_8)的最大电力。

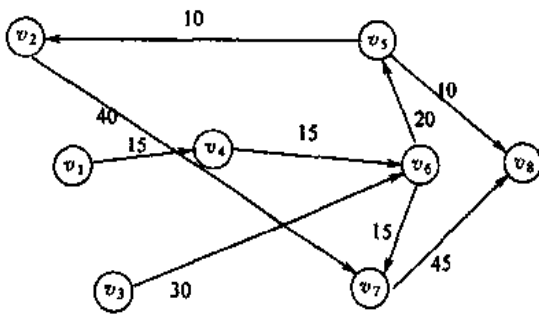


图 5.15

解 由于网络图只有一个发点和一个收点, 所以有必要添加一个虚设的起点 v_0 和弧 (v_0, v_1) 、 (v_0, v_2) 和 (v_0, v_3) 。弧上容量 $c_{01} = 15, c_{02} = 10, c_{03} = 40$ 分别为 v_1, v_2, v_3 三点的发电能力, 如图 5.16 所示。

第一步, 由观察法得一初始可行流 $\{f_{ij}\}$, 即图 5.16 上所标 (c_{ij}, f_{ij}) 。 c_{ij} 为弧 (v_i, v_j) 上容量, f_{ij} 为弧上流量。

第二步, 用标号法寻找可扩充链, 由此

检验当前可行流 $\{f_{ij}\}$ 是否是最大流, 标号过程如下。

①给 v_0 标号 $[\infty, v_0]$, 选与 v_0 关联的流出未饱和弧或流入非零流弧, 得流出未饱和弧 (v_0, v_3) , 给 v_3 标号 $[\theta_3, v_0]$, 其中 $\theta_3 = \min\{\infty, c_{03} - f_{03}\} = \min\{\infty, 40 - 10\} = 30$, 则 v_3 获得标号 $[30, v_0]$ 。

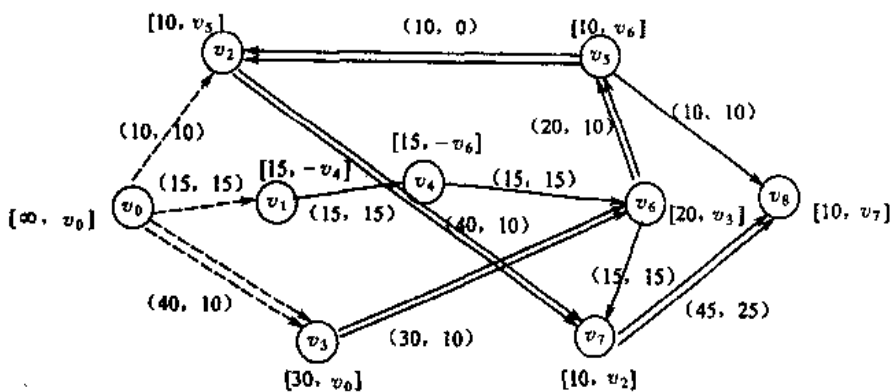


图 5.16

②把顶点集分为 $V_1 = \{v_0, v_3\}$, $\bar{V}_1 = \{v_1, v_2, v_4, v_5, v_6, v_7, v_8\}$, 考虑所有这样的弧 (v_i, v_j) 或 (v_j, v_i) 。其中 $v_i \in V_1, v_j \in \bar{V}_1$, 得弧 (v_0, v_2) 、 (v_0, v_1) 、 (v_3, v_6) 。在这些弧中, (v_3, v_6) 为流出未饱和弧, 给 v_6 标号 $[\theta_6, v_3]$, $\theta_6 = \min\{\theta_3 = 30, (30 - 10)\} = 20$, 即给 v_6 标号 $[20, v_3]$ 。

③把顶点集分为 $V_1 = \{v_0, v_3, v_6\}$, $\bar{V}_1 = \{v_1, v_2, v_4, v_5, v_7, v_8\}$, 考虑所有这样的弧 (v_i, v_j) 或 (v_j, v_i) 。其中 $v_i \in V_1, v_j \in \bar{V}_1$, 从中找出流出未饱和弧及流入非零流弧, 得流出未饱和弧 (v_6, v_5) 及流入非零流弧 (v_4, v_6) 。给 v_5 标号 $[\theta_5, v_6]$, 给 v_4 标号 $[\theta_4, -v_6]$ 。 $\theta_5 = \min\{\theta_6,$

$$c_{65} - f_{65} = \min\{20, 20 - 10\} = 10; \theta_4 = \min\{\theta_6, f_{46}\} = \min\{20, 15\} = 15。$$

④把顶点集分为 $V_1 = \{v_0, v_3, v_4, v_5, v_6\}$, $\bar{V}_1 = \{v_1, v_2, v_7, v_8\}$, 考虑所有这样的弧 (v_i, v_j) 或 (v_j, v_i) 。其中, $v_i \in V_1, v_j \in \bar{V}_1$, 从中找出流出未饱和弧及流入非零流弧。得流出未饱和弧 (v_5, v_2) , 则给 v_2 标号为 $[\theta_2, v_5]$, $\theta_2 = \min\{\theta_5, c_{52} - f_{52}\} = \min\{10, 10 - 0\} = 10$; 得流入非零流弧 (v_1, v_4) , 则给 v_1 标号为 $[\theta_1, -v_4]$, $\theta_1 = \min\{\theta_4, f_{14}\} = \min\{15, 15\} = 15。$

⑤把顶点集分为 $V_1 = \{v_0, v_1, v_2, v_3, v_4, v_5, v_6\}$, $\bar{V}_1 = \{v_7, v_8\}$, 考虑所有这样的弧 (v_i, v_j) 或 (v_j, v_i) 。其中 $v_i \in V_1, v_j \in \bar{V}_1$ 。从中找出流出未饱和弧及流入非零流弧, 得流出未饱和弧 (v_2, v_7) , 则给 v_7 标号 $[\theta_7, v_2]$ 。其中, $\theta_7 = \min\{\theta_2, c_{27} - f_{27}\} = \min\{10, 30\} = 10$, 则 v_7 获得标号 $[10, v_2]$ 。

⑥把顶点集分为 $V_1 = \{v_0, v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$, $\bar{V}_1 = \{v_8\}$ 。考虑弧 (v_i, v_j) 或 (v_j, v_i) , 其中, $v_i \in V_1, v_j \in \bar{V}_1$, 从中找出流出未饱和弧 (v_7, v_8) , 则给 v_8 标号 $[\theta_8, v_7]$, $\theta_8 = \min\{\theta_7, c_{78} - f_{78}\} = \min\{10, 45 - 25\} = 10$, 则 v_8 获得标号 $[10, v_7]$ 。

终点 v_8 已标号, 反向追踪可得 v_0 至 v_8 的可扩充链为

$$v_0 - v_3 - v_6 - v_5 - v_2 - v_7 - v_8$$

整个标号过程在图 5.16 上表示, 可扩充链用双线表示。

第三步, 调整流量。由所得可扩充链确定一新可行流 $\{\bar{f}_y\}$, 可扩充链上前向弧上各流量均增加 $\theta_7 = 10$, 即有

$$\begin{cases} \bar{f}_{03} = f_{03} + \theta_7 = 10 + 10 = 20 \\ \bar{f}_{36} = f_{36} + \theta_7 = 10 + 10 = 20 \\ \bar{f}_{65} = f_{65} + \theta_7 = 10 + 10 = 20 \\ \bar{f}_{52} = f_{52} + \theta_7 = 0 + 10 = 10 \\ \bar{f}_{27} = f_{27} + \theta_7 = 10 + 10 = 20 \\ \bar{f}_{78} = f_{78} + \theta_7 = 25 + 10 = 35 \end{cases}$$

链上无后向弧, 可扩充链各弧上流量不变, 新可行流 $\{\bar{f}_y\}$ 用图 5.17 表示。

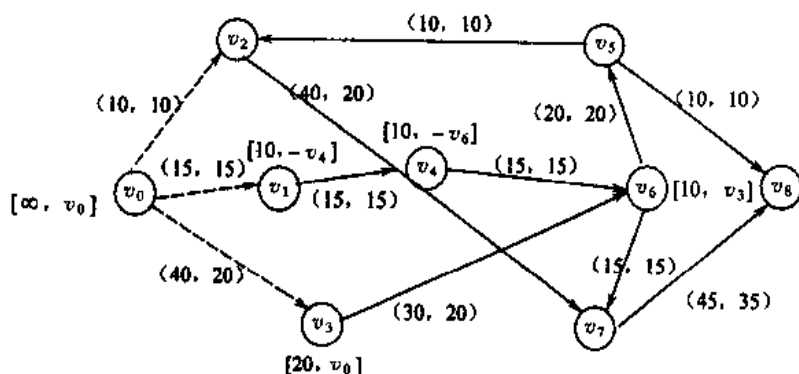


图 5.17

继续用标号法检验当前可行流是否为最大流, 标号过程如图 5.17 所示。

①给 v_0 标号 $[\infty, v_0]$, 选与 v_0 关联的流出未饱和弧或流入非零流弧, 得流出未饱和弧 (v_0, v_3) , 则给 v_3 标号 $[20, v_0]$ 。

②把顶点集分为 $V_1 = \{v_0, v_3\}$, $\bar{V}_1 = \{v_1, v_2, v_4, v_5, v_6, v_7, v_8\}$, 考虑弧 (v_i, v_j) 或 (v_j, v_i) , 其中 $v_i \in V_1, v_j \in \bar{V}_1$. 从中找出流出未饱和弧和流入非零流弧, 得流出未饱和弧 (v_3, v_6) , 则 v_6 获得标号 $[10, v_3]$ 。

③把顶点集分为 $V_1 = \{v_0, v_3, v_6\}$, $\bar{V}_1 = \{v_1, v_2, v_4, v_5, v_7, v_8\}$, 考虑弧 (v_i, v_j) 或 (v_j, v_i) , 其中 $v_i \in V_1, v_j \in \bar{V}_1$. 从中找出流出未饱和弧和流入非零流弧, 得流入非零流弧 (v_4, v_6) , 则 v_4 获得标号 $[10, -v_6]$ 。

④把顶点集分为 $V_1 = \{v_0, v_3, v_4, v_6\}$, $\bar{V}_1 = \{v_1, v_2, v_5, v_7, v_8\}$, 考虑弧 (v_i, v_j) 或 (v_j, v_i) , 其中 $v_i \in V_1, v_j \in \bar{V}_1$. 从中找出流出未饱和弧和流入非零流弧, 得流入非零流弧 (v_1, v_4) , 则 v_1 获得标号 $[10, -v_4]$ 。

⑤把顶点集分为 $V_1 = \{v_0, v_1, v_3, v_4, v_6\}$, $\bar{V}_1 = \{v_2, v_5, v_7, v_8\}$, 考虑弧 (v_i, v_j) 或 (v_j, v_i) , 其中 $v_i \in V_1, v_j \in \bar{V}_1$. 从中再找不到流出未饱和弧或流入非零流弧, 标号完毕, 且终点 v_8 未标号。这说明网络中已找不到可扩充链, 当前流即为最大流, 最大流流量为: $20 + 15 + 10 = 45$, 即三个发电站输送至 8 号地区的最大电力为 45 兆瓦。

进一步, 由第⑤步知, $V_1 = \{v_0, v_1, v_3, v_4, v_6\}$ 为已标号点集, $\bar{V}_1 = \{v_2, v_5, v_7, v_8\}$ 为未标号点集, 则最小截集由 V_1 流向 \bar{V}_1 的弧集构成。它是: $\{(v_0, v_2), (v_6, v_5), (v_6, v_7)\}$, 最小截集的截量为 $C(V_1, \bar{V}_1) = c_{02} + c_{65} + c_{67} = 10 + 20 + 15 = 45$ 。它恰好等于最大流的流量。

由上可见, 用标号法求解最大流问题, 同时得到一个最小截集。最小截集的容量的大小影响总的输送量的提高。因此, 为提高总的输送量, 必须首先考虑改善最小截集中各弧的输送状况, 提高它们的通过能力。因此, 在本例中, 若想进一步提高三个发电站输送到 8 号地区的电力的能力, 必须首先改善 6 号地区到 5 号地区和 7 号地区的电力输送能力。

习 题

5.1 求图 5.18 中各图的最小部分树。

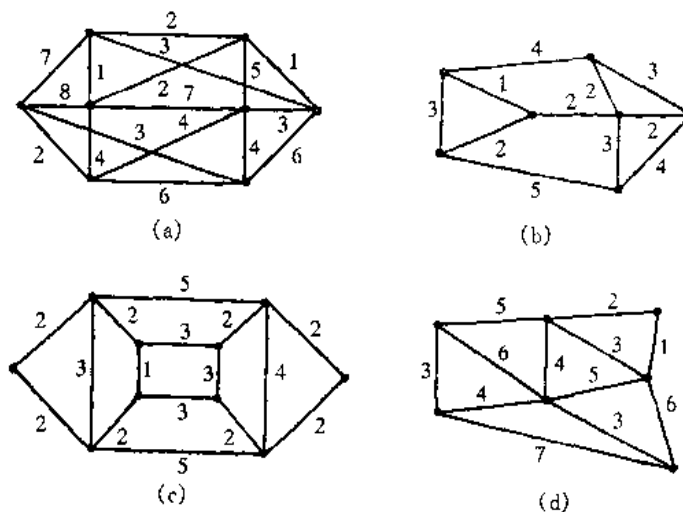


图 5.18

5.2 已知世界六大城市: (Pe)、(N)、(Pa)、(L)、(T)、(M), 试在由表 5.6 所示交通网络的数据中确定最小树。

表 5.6

	Pe	T	Pa	M	N	L
Pe	×	13	51	77	68	50
T	13	×	60	70	67	59
Pa	51	60	×	57	36	2
M	77	70	57	×	20	55
N	68	67	36	20	×	34
L	50	59	2	55	34	×

5.3 有 9 个城市 v_1, v_2, \dots, v_9 , 其公路网如图 5.19 所示。弧旁数字是该段公路的长度, 有一批货物要从 v_1 运到 v_9 , 问走哪条路最短。

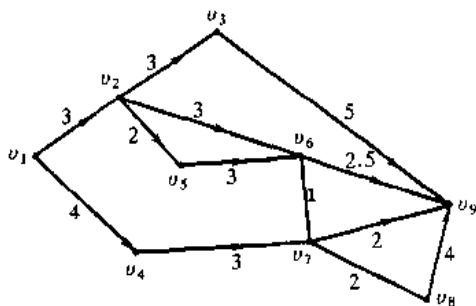


图 5.19

5.4 求如图 5.20 所示的网络的最大流和最小截集。(每弧旁的数字是 (c_{ij}, f_{ij}))。

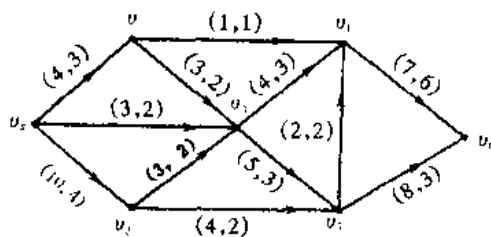
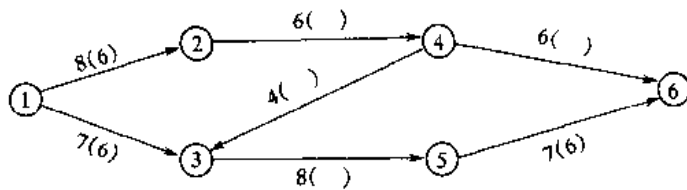


图 5.20

5.5 考虑如下图所示的网络最大流问题, 其中弧上的数字为容量, 括号内的数字为流量。

(1) 在括号内填上适当的数字, 使构成一个可行流。



(2) 若选择点集 V_1 和 \bar{V}_1 如表 5.7, 请填出相应的截集与截量。

表 5.7

V_1	\bar{V}_1	(V_1, \bar{V}_1)	$C(V_1, \bar{V}_1)$
①, ③	②, ④, ⑤, ⑥		

5.6 某汽车公司正在制订 5 年内购买汽车的计划。下面给出一辆新汽车的价格(表 5.8)以及一辆汽车的使用维修费用(万元)(表 5.9)。试用网络分析中求最短路的方法确定公司可采用的最优策略。

表 5.8

年号	1	2	3	4	5
价格	2	2.1	2.3	2.4	2.6

表 5.9

汽车使用年龄	0~1	1~2	2~3	3~4	4~5
维修费用	0.7	1.1	1.5	2	2.5

第6章 网络计划

6.1 网络计划图的绘制

从20世纪50年代起,人们就试用网络方法来编制计划,用网络来代替传统的横道图,其中最著名的是“关键路径法”(Critical Path Method,简称CPM方法)和“计划评审技术”(Program Evaluation and Review Technique,简称PERT法)。PERT方法是美国海军特种工程处在波兹(BOOZ)、阿伦(Allen)和汉密尔顿(Hamilton)管理咨询公司协助下首先研究出来的。这种网络方法因在美国北极星导弹研制工程计划中成功的应用而著名。CPM方法是1957年由美国兰德公司的J.E.凯雷和杜邦公司的M.R.瓦尔科提出,并首次应用于一个化学工厂设备维修工作计划中。

CPM与PERT网络计划技术的原理基本一致,都是用网络表示工程项目,以确定关键路线。这一思想是它们的基础。这两种方法的主要区别是:

①CPM方法对工程中各工序完工时间的估计是确定的,而PERT方法对时间估计是概率型的;

②PERT方法主要注重工程计划中的时间问题,而CPM方法在把费用概念引入计划控制过程中,不仅注重时间问题,而且还考虑工程的费用和成本问题,考虑工程时间的均衡问题。

在实际工作中,若一个工程的工序时间比较确定,费用也能预先估算得比较准确(例如一般的建筑工程),使用CPM方法比PERT更为优越,而当一个工程中不确定因素程度较多时,控制工程费用选PERT方法更为合适。

近30年的实践表明,网络方法是一种非常有效的特殊计算方法。这种方法通过网络图及相应的分析计算,能够使计划工作统筹兼顾,全面安排,并能指示在执行计划中对全局有重大影响的关键所在。

以下较详细地介绍网络计划方法。

网络计划方法一般由两个阶段组成:初始计划阶段和计划方案优化调整阶段。前者是先把工程划分为多个相互衔接的工序,估计工序完工时间,然后用网络图表示,按计划时间参数确定完成工程的各关键工序及由其组成的关键路线,拟定一初始计划方案。后者则根据要求,综合考虑时间、费用、咨询等目标,对初始计划方案进行调整改善,直至得出一满意的计划方案。其中,前一初始计划阶段又包括网络绘制与关键路线确定两个环节。本节首先介绍网络图的绘制。

6.1.1 网络图(箭头图)的组成要素

一项工程总由许多彼此关联的独立活动组成。这些活动称为工序。各道工序之间先后关联,完成每道工序的时间称为工序时间。因此,可以采用一个网络图(赋权的有向图)表示各道工序之间的相互依存的顺序及工序时间。

一般表示工程计划的网络图由三部分组成,即箭线、结点和线路。

1)实箭线 它表示一道具体工序。箭头表示工序进行方向,箭尾表示工序开始,箭头表示

工序结束。箭头和箭尾与结点相连,结点用圆圈表示,并编上号码。箭线二端的结点号码就表示一道工序,通常称为双代号表示法。

2)虚箭线 它表示一道虚工序。虚工序不是实际中的具体工序,它仅用于表示工序与工序之间的关联关系。虚工序不需时间、费用和资源。

3)结点 它表示一个事项。事项又称事件,表示一些工序的结束或开始。在相邻工序首尾衔接处的结点用圆圈表示,圈内注上该结点(事项)的序号。连接箭尾的结点称为该工序的紧前事项,连接箭头的结点称为该工序的紧后事项。整个网络图起始结点称为总开工事项,而最后工序的结束结点为完工事项。网络中所有工序都是用事项(结点)进行相互联系并互相制约的,只有当事项的所有紧前工序完成后,事项的紧后工序才能开始。

4)线路 它是指从最初结点(总开工事项)开始顺着箭头方向连续不断地到达终点(总完工事项)的通路,这样通路有许多条,各条线路上工序时间的和可能不同,其中时间和最大的线路为关键线路,相应有关键线路。关键线路上的工序称为关键工序。

6.1.2 绘制网络图的准备工作

在绘制工程计划网络图前,首先要完成下列准备工作。

1. 确定目标

工程计划目标是多方面综合的,但按侧重点不同大致可分成三类:第一类,时间要求为主;第二类,资源要求为主;第三类,费用要求为主。

2. 工程任务的分解和分析

分析工程由哪些工序组成,并列出全部工序及其代号清单。工程任务的分解,根据不同对象有不同要求。例如,对领导机构,重要的是纵观全局,掌握关键,因此,工程可以分解得粗一些;对基层生产单位,要根据网络图组织生产,因此,工程应该分解得细一些。

3. 确定各工序之间先后顺序及衔接关系

要确定每一道工序开工之前有哪些工序必须先期完成,即确定每一道工序的紧前工序是哪些。

4. 确定各工序时间

要确定每道工序的完成所需时间,称为工序时间。

后三项工作经常用工序清单的形式给出。例如,某工厂进行技术改造,需要拆掉旧厂房、建造新厂房和安排设备。这项改建工程可以分解为7道工序:拆迁(A)、工程设计(B)、土建工程设计(C)、采购设备(D)、厂房土建(E)、设备安装(F)、设备调试(G)。它们的顺序关系用表6.1表示。

表 6.1

工序代号	工序名称	紧前工序	工序时间(周)
A	拆迁	\	2
B	工程设计	\	3
C	土建工程设计	B	2.5
D	采购设备	B	6
E	厂房土建	C, A	20
F	设备安装	D, E	4
G	设备调试	F	2

6.1.3 网络图绘制规则

网络图是有向的,绘制时从左开始向右延伸。网络图中不能出现回路,即不可有循环现象,否则出现逻辑错误,工作永远也达不到终点。图 6.1 所示情况是不允许的。

二个事项之间只能有一个工序,即二条箭线不能有同样的始末点。若二个事项之间有几个平行进行的工序,不许直接联起来,必须引入虚工序。如图 6.2(a)是错误的,正确的画法如图 6.2(b)所示。

为正确表示工序之间的衔接关系,网络图中引进了“虚工序”概念。如图 6.3 所示,虚工序 E 只表示 D 工序在 B、C 工序完工后才能开工。

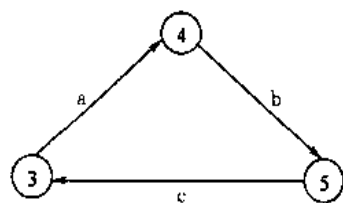


图 6.1

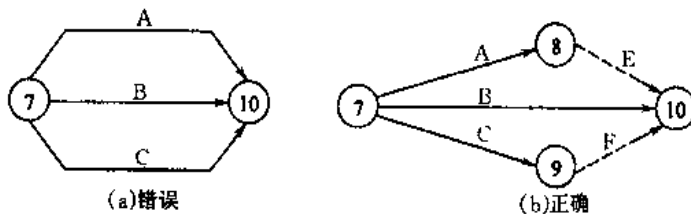


图 6.2

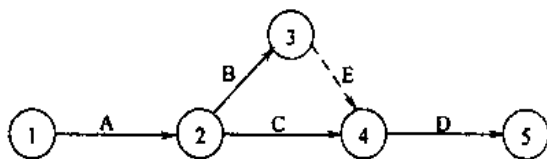


图 6.3

平行作业就是指两项以上的工序从同一紧前事项引出,又有同样的紧后事项。为了避免二个事项之间有两条以上箭线相连,必须引入虚工序,如图 6.4 所示。修理(1)、修理(2)和修理(3)三道工序可以在拆卸工序后平行作业,添加了虚工序 E、F,以便正确表示这种平行作业。

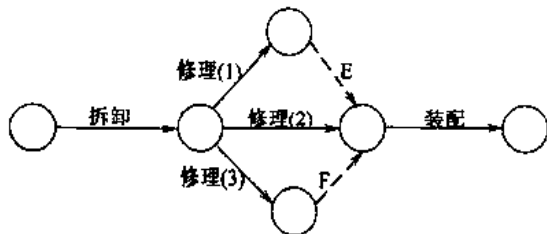


图 6.4

交叉作业指一项工作不必全部完工后才开始下一道工序,而是前道工序完成一部分,就开始后道工序,待前道工序再完成一部分,后道工序也完成一部分并接着继续做下一部分,这样形成工序之间一部分一部分的交叉进行。对交叉作业也需引入虚工序表示。例如图 6.5 中, A 和 B 分别表示工装设计和工装制造二道工序,可以一部分设计一部分制造交叉进行,以便缩

短总工期。设 $A = a_1 + a_2 + a_3$, $B = b_1 + b_2 + b_3$, 引入虚工序则能正确表示这种交叉作业。

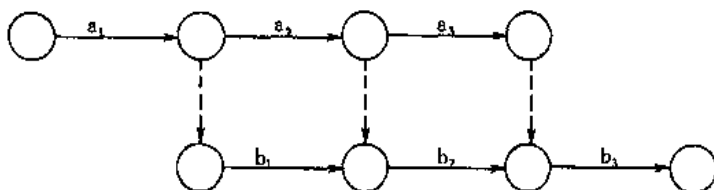


图 6.5

网络图中一般只有一个总开工事项和一个总完工事项。若一个工程的网络图同时有几个起始点或几个终点时,为了使网络图更加清晰,经常在始点和终点添加一些虚工序,使网络图上只有一个起点和终点,如图 6.6 所示。

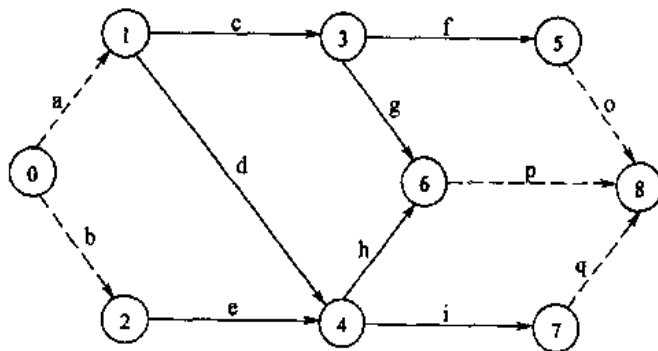


图 6.6

网络图中的事项要统一编号。编号由左至右进行,每道工序的箭尾事项号应小于箭头事项号。编号的具体方法如下(箭杆删除法):首先给起点(总开工事项)以编号 1,接着设想把从起点流出箭线删除,得到一个或几个不为任何有向边终点的结点(不为任何箭线箭头事项的结点),对它们逐一编上尽可能小的号(这些结点之间编号大小的顺序是无关紧要的),然后把把这些新编号结点流出的箭线删除,又得一些不为任何箭线终点的结点,同样逐一编号,如此继续进行,直至最后一个结点(即总完工事项)被编上号为止。参看图 6.7 网络编号过程。

6.1.4 确定每道工序完成时间

绘制一张网络图除要注意上述规则外,还必须用合适的方法估计每道工序完成时间,即确定工序时间 $t(i, j)$ 。这里 i 是工序箭尾结点号, j 是工序箭头结点号。通常确定工序时间 $t(i, j)$ 有二种方法。

1. 一时估计法

此法又称单一时间估计法,即对各道工序的完成时间仅估计一个数值。估计时,以完成工序所需要最大作业时间值为准。

在工序的工时定额资料比较健全或者有同类工序时间的统计资料时,经常使用一时估计法,根据资料,通过分析对比确定各工序的工序时间。

2. 三时估计法

对工序的作业时间,预先估计三个时间,然后求出可能完成的平均值。三个估计时间值为: a 为最乐观时间,指顺利完成的最短时间; b 为最悲观时间,指极不顺利条件下完成工序的

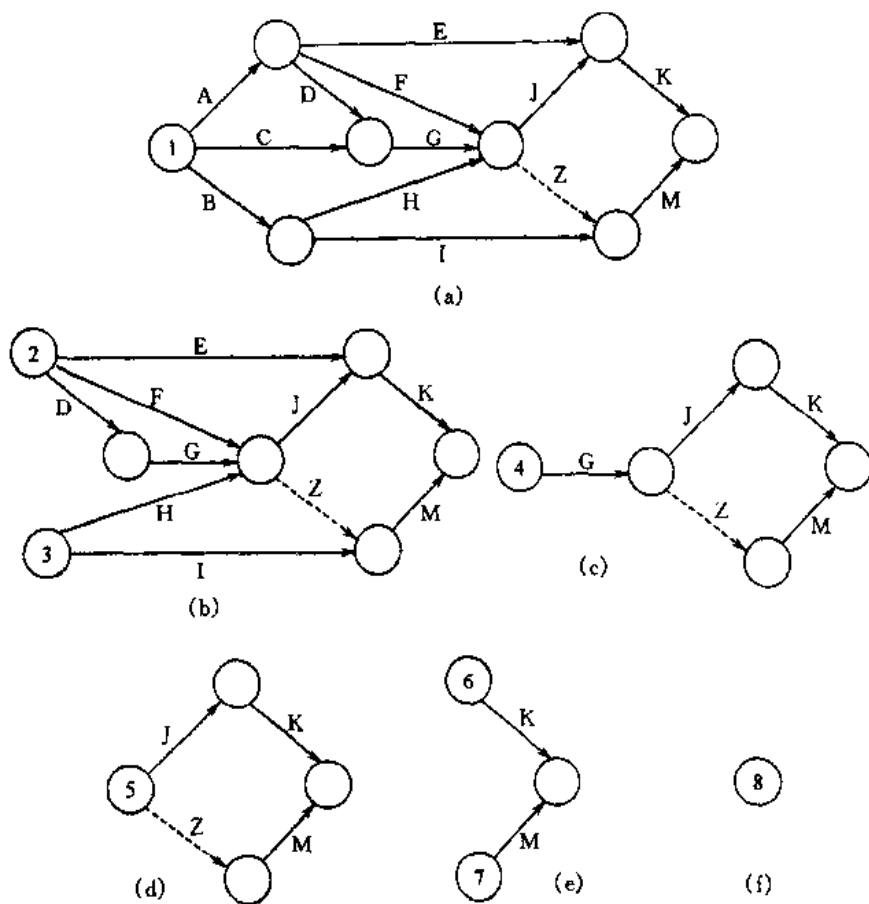


图 6.7 事项编号过程

最长时间; m 为最可能时间, 指正常情况下完成工序最可能的时间。

工序时间 t 用下列公式来确定:

$$t = \frac{a + 4m + b}{6}$$

以下通过例子说明网络图的绘制过程。

例 6.1 根据表 6.1 给出的某厂技术改造工程施工情单, 绘制该工程的逻辑网络图如下:

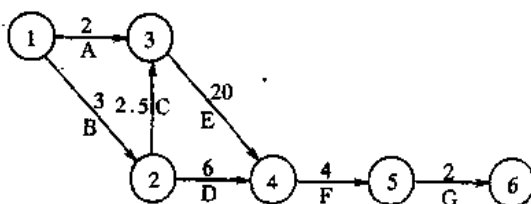


图 6.8

例 6.2 某化工厂拟进行管道安装工程, 工程的组成及工艺流程已由表 6.2 给出。绘制出此工程相应的网络图(图 6.9)。

表 6.2

活动代号	活动项目	紧前工作	时间(h)
A	器材调查	/	8
B	停止旧管道工作	A	8
C	搭脚手架	A	12
D	拆除旧管道阀门	B,C	35
E	准备阀门	A	225
F	准备管道	A	200
G	组装一部分管道	F	40
H	安装新管道	D,E,G	32
I	安装阀门	D,E	8
J	焊接管道	H,I	8
K	装配管道和阀门	J	8
L	包扎管道	J	24
M	拆脚手架	K,L	4
N	压力试验	K	6
P	整理现场	M,N	4

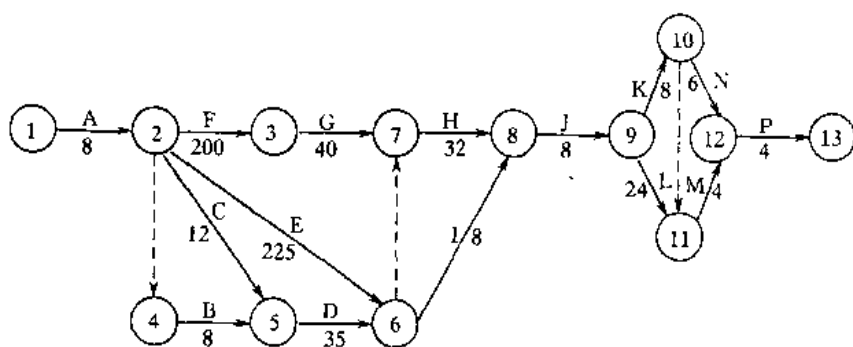


图 6.9

6.2 时间参数计算与关键路线确定

6.2.1 时间参数计算

网络图上的时间参数主要包括事项最早时间、事项最迟时间等几项时间指标。

1. 事项最早时间 $t_E(i)$

这是指事项之最早可能发生时间。其计算方法是从始点事项开始,设 $t_E(1)=0$,表示工程从零时刻开工,然后自左至右逐步计算各事项最早时间,直至终点事项。一个箭头事项的最早时间等于相关箭尾事项的最早时间加上该箭线时间(即工序时间)。若一个事项同时是几个箭线的箭头事项,则选择其中箭尾事项的最早时间加箭线时间之和的最大值作为该事项最早时间。这可用下列公式确定:

$$\begin{cases} t_E(1) = 0 \\ t_E(j) = \max\{t_E(i) + t(i, j)\} \quad (j = 2, 3, \dots, n) \end{cases}$$

终点事项最早时间记为 $T_E = t_E(n)$, 通常称 T_E 为工程的最早完工期, 简称为工程工期。从网络角度看, T_E 就是从起点到终点最长路的路长。其中箭线的路长表示工序时间。

事项最早时间 $t_E(j)$ 的计算方法归纳如下:

- ① $t_E(1) = 0$;
- ② 从左至右计算;
- ③ $t_E(j) = \max\{t_E(i) + t(i, j)\} \quad (j = 2, 3, \dots, n)$;
- ④ $T_E = t_E(n)$ 。

例如, 图 6.10 中, 已知 $t_E(7) = 5, t_E(8) = 6$, 计算事项最早时间为

$$\begin{aligned} t_E(9) &= \max\{[t_E(7) + t(7, 9)], [t_E(8) + t(8, 9)]\} \\ &= \max\{(5 + 6), (6 + 7)\} = 13 \end{aligned}$$

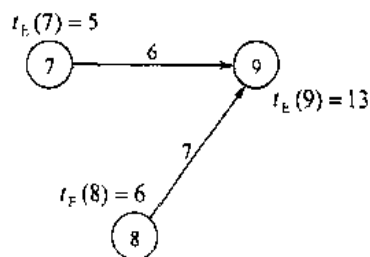


图 6.10

2. 事项最迟时间 $t_L(i)$

一个事项若晚于某一时刻发生, 就会推迟整个工程的最早完工期, 这个时间称为事项最迟时间。终点事项(总完工事项)最迟时间就是工程最早完工工期, 即有 $t_L(n) = T_E$ 。事项最迟时间是从终点事项起, 自右至左逐个计算, 即有公式:

$$\begin{cases} t_L(n) = t_E = t_E(n) \\ t_L(i) = \min\{t_L(j) - t(i, j)\} \end{cases}$$

事项最迟时间计算方法归纳如下:

- ① $t_L(n) = T_E = t_E(n)$;
- ② 从右往左计算;
- ③ $t_L(i) = \min\{t_L(j) - t(i, j)\} \quad (i = n - 1, n - 2, \dots, 1)$ 。

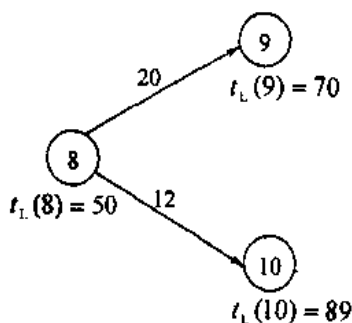


图 6.11

如图 6.11 所示, 若已算出事项 9 和事项 10 的最迟时间为 $t_L(9) = 70, t_L(10) = 89$, 则计算出事项 8 的最迟时间为:

$$\begin{aligned} t_L(8) &= \min\{[t_L(9) - t(8, 9)], [t_L(10) - t(8, 10)]\} \\ &= \min\{(70 - 20), (89 - 12)\} \\ &= 50 \end{aligned}$$

3. 工序 (i, j) 最早可能开工时间 $t_{ES}(i, j)$ (earliest start time)

一道工序必须在其所有紧前工序完工后才能开工, 所以工序 (i, j) 最早可能开工时间即为工序箭尾事项的最早时间, 即有

$$t_{ES}(i, j) = t_E(i)$$

4. 工序 (i, j) 最迟必须开工时间 $t_{LS}(i, j)$ (latest start time)

这个参数是指在不影响整个工期 T_E 的条件下工序最迟必须开始的时刻, 等于这个工序

箭头事项最迟时间减去工序时间,即有

$$t_{LS}(i, j) = t_L(j) - t(i, j)$$

5. 工序 (i, j) 最早可能完工时间 $t_{EF}(i, j)$ (earliest finish time) 和最迟必须完工时间 $t_{LF}(i, j)$ (latest finish time)

显然,它们分别为

$$t_{EF}(i, j) = t_{ES}(i, j) + t(i, j)$$

$$t_{LF}(i, j) = t_{LS}(i, j) + t(i, j)$$

6. 工序 (i, j) 的总时差 $R(i, j)$

在不影响整个工程工期 T_E 的条件下,工序最早可能开工时间可以推迟的时间数称为工序的总时差,它表示工序安排上可以松动的时间数。显然,总时差为

$$R(i, j) = t_{LS}(i, j) - t_{ES}(i, j) = t_{LF}(i, j) - t_{EF}(i, j)$$

7. 工序 (i, j) 的单时差 $r(i, j)$

单时差是指不影响紧后工序最早可能开工时间条件下,工序最早可能完工时间可以推迟的时间,即

$$r(i, j) = t_E(j) - t_{EF}(i, j)$$

6.2.2 确定关键路线

根据网络图中的参数,可以确定关键路线。

要确定关键路线,必须先找出关键工序。关键工序是指总时差为零的工序。关键路线是由关键工序连接而成的线路。

从网络角度看,关键路线就是从起点至终点的最长路(箭线路长表示工序时间),它决定整个工程工期 T_E 的长短。确定关键路线是网络计划方法中重要的环节。关键路线不惟一,计划安排得越紧凑,关键路线越多。

计算网络图的时间参数和确定关键路线的方法有图上作业法、表格计算和矩阵计算方法等。图上作业法用得最多,但对一些复杂大型工程的网络则必须用计算机计算。

例 6.3 某项课题科研工作分解的作业表由表 6.3 给出,根据此表绘制此项科研工作的网络图,计算时间参数,并确定关键路线。

表 6.3

工序代号	工 序	紧前工序	工序时间
A	系统提出和研究问题	无	4
B	研究选点问题	A	7
C	准备调研方案	A	10
D	收集资料工作安排	B	8
E	挑选、实地训练工作人员	B, C	12
F	准备收集资料用表格	C	7
G	实地调查	D, E, F	5
H	分析准备调查报告	G	4
I	任务结束	H	0

解 第一步,按规则绘制网络图(图 6.12)。

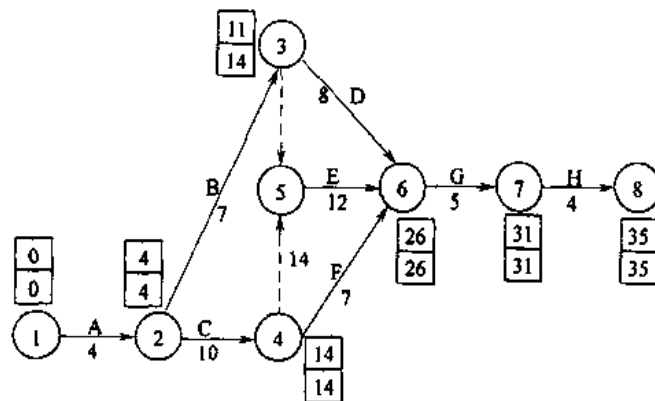


图 6.12

第二步,计算各时间参数。步骤如下:

①从左至右逐步计算事项最早时间 $t_E(i)$,得

$$t_E(1) = 0$$

$$t_E(2) = t_E(1) + t(1,2) = 0 + 4 = 4$$

$$t_E(3) = t_E(2) + t(2,3) = 4 + 7 = 11$$

$$t_E(4) = t_E(2) + t(2,4) = 4 + 10 = 14$$

$$t_E(5) = \max \begin{cases} t_E(3) + t(3,5) \\ t_E(4) + t(4,5) \end{cases} = \max \begin{cases} 11 + 0 \\ 14 + 0 \end{cases} = 14$$

$$t_E(6) = \max \begin{cases} t_E(3) + t(3,6) \\ t_E(5) + t(5,6) \\ t_E(4) + t(4,6) \end{cases} = \max \begin{cases} 11 + 8 \\ 14 + 12 \\ 14 + 7 \end{cases} = 26$$

$$t_E(7) = t_E(6) + t(6,7) = 26 + 5 = 31$$

$$t_E(8) = t_E(7) + t(7,8) = 31 + 4 = 35$$

整个工程最早完工期 $T_E = t_E(8) = 35$ 。

②从右到左逐步计算事项最迟时间 $t_L(i)$,得

$$t_L(8) = T_E = 35$$

$$t_L(7) = t_L(8) - t(7,8) = 35 - 4 = 31$$

$$t_L(6) = t_L(7) - t(6,7) = 31 - 5 = 26$$

$$t_L(5) = t_L(6) - t(5,6) = 26 - 12 = 14$$

$$t_L(4) = \min \begin{cases} t_L(6) - t(4,6) \\ t_L(5) - t(4,5) \end{cases} = \min \begin{cases} 26 - 7 \\ 14 - 0 \end{cases} = 14$$

$$t_L(3) = \min \begin{cases} t_L(6) - t(3,6) \\ t_L(5) - t(3,5) \end{cases} = \min \begin{cases} 26 - 8 \\ 14 - 0 \end{cases} = 14$$

$$t_L(2) = \min \begin{cases} t_L(3) - t(2,3) \\ t_L(4) - t(2,4) \end{cases} = \min \begin{cases} 14 - 7 \\ 14 - 10 \end{cases} = 4$$

$$t_L(1) = t_L(2) - t(1,2) = 4 - 4 = 0$$

将已求得的 $t_E(i)$ 、 $t_L(i)$ 分别填入网络图事项 i 边上的上、下小方格内, 整个计算过程可以在图上进行。

③计算各工序最早开工时间和最迟必须开工时间:

$$t_{ES}(i,j) = t_E(i)$$

$$t_{ES}(1,2) = t_E(1) = 0, t_{ES}(2,3) = t_E(2) = 4$$

$$t_{ES}(2,4) = t_E(2) = 4, t_{ES}(3,6) = t_E(3) = 11$$

$$t_{ES}(5,6) = t_E(5) = 14, t_{ES}(4,6) = t_E(4) = 14$$

$$t_{ES}(6,7) = t_E(6) = 26, t_{ES}(7,8) = t_E(7) = 31$$

$$t_{LS}(i,j) = t_L(j) - t(i,j)$$

$$t_{LS}(1,2) = t_L(2) - t(1,2) = 4 - 4 = 0$$

$$t_{LS}(2,3) = t_L(3) - t(2,3) = 14 - 7 = 7$$

$$t_{LS}(2,4) = t_L(4) - t(2,4) = 14 - 10 = 4$$

$$t_{LS}(3,6) = t_L(6) - t(3,6) = 26 - 8 = 18$$

$$t_{LS}(5,6) = t_L(6) - t(5,6) = 26 - 12 = 14$$

$$t_{LS}(4,6) = t_L(6) - t(4,6) = 26 - 7 = 19$$

$$t_{LS}(6,7) = t_L(7) - t(6,7) = 31 - 5 = 26$$

$$t_{LS}(7,8) = t_L(8) - t(7,8) = 35 - 4 = 31$$

④计算各工序的总时差和单时差:

$$R(i,j) = t_{LS}(i,j) - t_{ES}(i,j)$$

$$R(1,2) = 0 - 0 = 0, R(2,3) = 7 - 4 = 3$$

$$R(2,4) = 4 - 4 = 0, R(3,6) = 18 - 11 = 7$$

$$R(5,6) = 14 - 14 = 0, R(4,6) = 19 - 14 = 5$$

$$R(6,7) = 26 - 26 = 0, R(7,8) = 31 - 31 = 0$$

$$r(i,j) = t_E(j) - t_{EF}(i,j) = t_E(j) - t_E(i) - t(i,j)$$

$$r(1,2) = 4 - 0 - 4 = 0, r(2,3) = 11 - 4 - 7 = 0$$

$$r(2,4) = 14 - 4 - 10 = 0, r(3,6) = 26 - 11 - 8 = 7$$

$$r(4,6) = 26 - 14 - 7 = 5, r(6,7) = 31 - 26 - 5 = 0$$

$$r(5,6) = 26 - 14 - 12 = 0, r(7,8) = 35 - 31 - 4 = 0$$

上述计算所得各参数分别列入表 6.4 中。

表 6.4

工序(i,j)	工序时间 $t(i,j)$	最早开工时间 $t_{ES}(i,j)$	最迟开工时间 $t_{LS}(i,j)$	总时差 $R(i,j)$	单时差 $r(i,j)$	关键工序
A(1,2)	4	0	0	0	0	×
B(2,3)	7	4	7	3	0	
C(2,4)	10	4	4	0	0	×
D(3,6)	8	11	18	7	7	

续表

工序(i, j)	工序时间 $t(i, j)$	最早开工时间 $t_{ES}(i, j)$	最迟开工时间 $t_{LS}(i, j)$	总时差 $R(i, j)$	单时差 $r(i, j)$	关键工序
E(5,6)	12	14	14	0	0	×
F(4,6)	7	14	19	5	5	
G(6,7)	5	26	26	0	0	×
H(7,8)	4	31	31	0	0	×

总时差为零的工序为关键工序,如 A(1,2)、C(2,4)、E(5,6)、G(6,7)、H(7,8)。关键路线为

①→②→④→⑤→⑥→⑦→⑧

关键路线在网络上用粗线或双线表示。

6.3 网络图的调整及优化

6.3.1 缩短工程工期问题

缩短工程工期是网络计划方法要解决的主要问题之一,当然,应该在保证质量和不增加人力、物力的前提下,尽量缩短工期。具体做法大致可概括为:在工程网络图上确定关键路线,明确哪些工序是关键工序,然后设法缩短关键工序的时间,从而达到缩短工期的目的。经常采用的措施如下。

1)压缩关键工序的工序时间 在关键工序上采取改进技术、工艺和设备等措施,尽先保证关键工序所需的人力和物力。当非关键工序和关键工序在人力、物力上有矛盾时,非关键工序要尽可能让路,以图缩短关键工序时间。

2)在非关键工序上尽量挖掘潜力 利用非关键工序的时差(机动余地)进行合理调度,抽调人力、物力支援关键工序,缩短关键工序时间。例如,造船工程计划中,下料是关键工序,而装配不是关键工序,且有时差(即有机动时间),因此把一部分装配工人调来突击下料。

3)尽量采用平行作业和交叉作业 如果能把一道工序拆成几道平行进行,无疑可以缩短作业时间,例如挖沟,两头同时挖比一头挖快。对相连的几道工序尽量交叉进行,例如在机床检修中,拆卸、清洗和检查这三道工序也可交叉进行,也就是拆一部分就洗,洗一部分就检查,这样可以缩短这三道工序的总时间。

4)注意关键路线的变化 在计划执行过程中,要特别注意关键路线的变化。在实施过程中,工程计划网络图决不会一成不变。因此,在计划执行过程中,由于压缩关键工序时间,可能会引起关键路线变化,必须充分注意生产进展情况,注意关键路线的变化,并相应地调整网络图。

6.3.2 工程的时间费用分析

编制工程网络计划时,不仅要考虑工期时间,还需考虑工程造价,因此,需计算工程在不同完工期时所需的费用。对一个工程来说,无论是以考虑工程费用为主,还是以控制时间进度为主,都必须对工程计划进行时间—费用分析,从而可综合考虑制定最优的计划。

1. 有关费用的一些概念

工程的成本费用由直接费用和间接费用两部分组成。

1) 直接费用 这是指人工、材料、燃料等直接用来完成工程任务的费用。

2) 间接费用 这是指管理人员的工资、办公、采购等并非直接用于完成工程任务的费用。

3) 工序费用率 这是指一道工序缩短单位时间所需增加的费用,可用公式表示:

$$\text{费用率 } q = \frac{\text{赶进度工时费用} - \text{正常工序时间费用}}{\text{正常工序时间} - \text{赶进度工序时间}}$$

一般来讲,缩短工期会增加直接费用,减少间接费用,延长工期引起直接费用减少和间接费用增加。在进行时间—费用分析时,要确定一个工程成本(直接费用和间接费用之和)最低的工期,通常称为最佳工期。寻找最佳工期通常使用所谓 CPM——成本法。为了使问题简单化,一般假设工序的直接费用与工序时间成反比,而工序的间接费用与工序时间成正比,如图 6.13 所示。

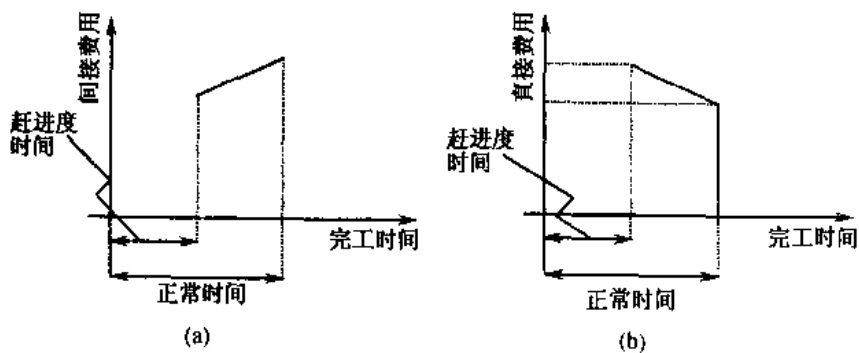


图 6.13

2. 工程的时间—费用关系图

在网络计划图上通过时间参数计算能确定工程期 T_E , 通常称为正常情况下的工期。而在进行工程时间—费用分析时,经常遇到“赶工期”问题,即分析工程由正常工期经赶工期后(缩短工期),工程费用如何变化? 讨论经缩短后各完工期与相应工程费用之间的关系,从而去寻求最低成本工期,这种分析经常用工程时间—费用关系图描述。图 6.14 就是典型工程时间—费用关系图。

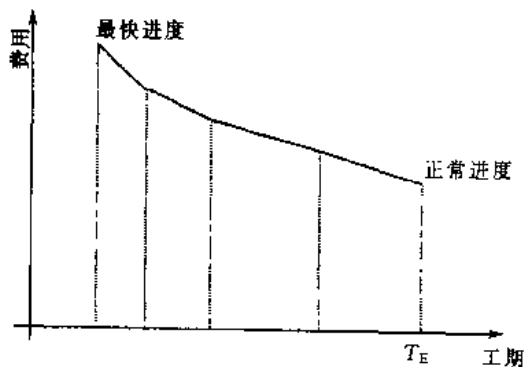


图 6.14

对工程的时间—费用关系图,要注意以下几点。

① 工程的时间—费用曲线,是所有的费用—延续时间点的下界包络线,这些点可通过压缩(缩短)网络中工序得到,例如,若缩短非关键工序,工程总费用增加,但工程总工期保持不变。

② 赶工期问题中,只有压缩关键工序才能引起总工期缩短。

③ 若网络图中同时有几条关键路线,为了缩短总工期,则几条关键路线必须同时在各自线路上压缩一个关键工序,否则,一条或几条线路可能变成非关键路线,这样并没能缩短总工期,却增加了工程费用。

④ 由于工期的时间—费用曲线是所有可能工期的费用曲线之下界包络,因而在该曲线上每个工期对应的费用就代表相应工期内完成该工程的最低直接费用。

上述工程的时间—费用曲线是一条直接费用曲线。为了确定赶工期问题中的最低成本工期,可将间接费用并入关系图,把同一工期下的直接费用和间接费用相加,就得到总费用曲线,从总费用曲线上能确定最低成本工期。这个做法可用图 6.15 说明。

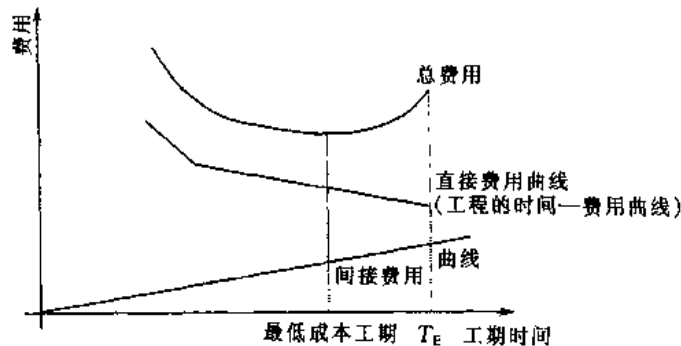


图 6.15

3. 赶工期问题中最低成本工期的确定

为了要确定一个工程的最低成本工期,首先要根据已有资料确定以下参数:

$q(i, j)$ ——每道工序 (i, j) 的费用率;

$\bar{t}(i, j)$ ——每道工序 (i, j) 上,工序时间缩短的下限;

p ——单位时间的间接费用。

确定最低成本工期的步骤如下:

①绘制网络图,计算工序总时差 $R(i, j)$,确定 T_E 与关键路线,计算相应工程费用;

②在各条关键路线上所有可压缩的工序中,各确定一个各自线路上费用率最低的工序作为压缩工序,这些压缩工序之集合用 I 表示;

③计算这些压缩工序上费用率之和 $\sum_{(i, j) \in I} q(i, j) = q$,若 $(q - p) > 0$,则表示压缩工期将引起工程费用上升,若 $(q - p) < 0$,表示可以压缩工期使总费用下降;

④确定压缩时间 t , t 的选择可由 $\bar{t}(i, j)$ 与 $R(i, j)$ 决定,即 t 取压缩工序上 $[t(i, j) - \bar{t}(i, j)]$ 与一切非零总时差之中最小者,可用下列式子表示:

$$t = \min\{\alpha, \beta\}$$

$$\alpha = \min_{(i, j) \in I} \{t(i, j) - \bar{t}(i, j)\}$$

$$\beta = \min\{R(i, j) | R(i, j) \neq 0\}$$

在压缩工序上压缩时间 t 以后,绘制新的网络图,继续重复上述步骤,直到再不能压缩为止。

下面举例说明网络计划中,使用时间与费用均衡的方法确定最低成本工期的步骤。

例 6.4 图 6.16 表示由 l、m、n、o、p、q、r 七道工序组成的网络图及有关时间参数,工程有关费用由表 6.5 给出,正常工期 $T_E = 60$ 天,试确定赶工期问题中最低成本工期。

表 6.5

工序	正常完工进度直接费用(元)	各工序费用率
l	800	/
m	3 000	180
n	2 000	120

续表

工序	正常完工进度直接费用(元)	各工序费用率
o	1 500	5 天以内为 60 6 天以上为 100
p	3 000	/
q	500	120
r	2 000	150
合计	12 800	/
间接费用	每天为 200 元	

解 由给出资料可确定各工序的费用率(表 6.5).各工序压缩时间下限 $\bar{t}(i, j) = 0$, 每天间接费用 $p = 200$ 元。

[循环 1]

第一步, 建立网络图(图 6.16), 确定各工序总时差 $R(i, j)$, 见表 6.6。关键路线由 q 和 r 二道工序组成, 工序时间 $T_E = 60$ 天。

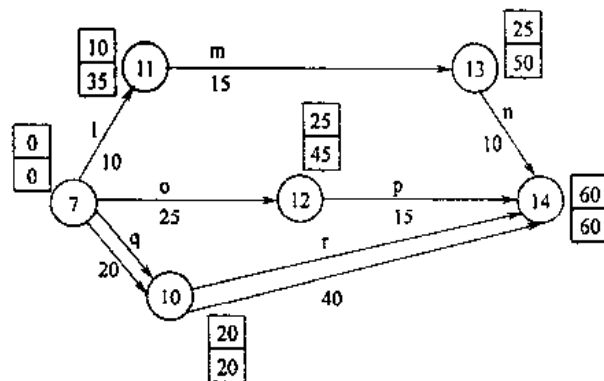


图 6.16

方案 1: 工期为 60 天(正常工期)的费用: $S_1 = 12\,800 + 200 \times 60 = 24\,800$ 元

第二步, 关键路线上只能缩短 r 工序, 则 r 工序为压缩工序。

表 6.6

工序	总时差 $R(i, j)$	工序	总时差 $R(i, j)$
l	25	p	20
m	25	q	0
n	25	r	0
o	20		

第三步, r 工序费用率 $q_r = 150$, $(q_r - p) = -50 < 0$ 。可以进一步压缩。

第四步, 确定压缩时间 t 。显然有 $\alpha = 40$, $\beta = 20$, $t = 20$, 在 r 工序上压缩时间 20 天, 得新工序时间。

[循环 2]

第一步, 建立网络图(图 6.17), 确定各工序上总时差 $R(i, j)$, 见表 6.7, 有二条关键路线 $(7) \rightarrow (10) \rightarrow (14)$ 和 $(7) \rightarrow (12) \rightarrow (14)$, 完工期为 40 天。

方案 2:工期为 40 天时的工程费用为

$$S_2 = 12\,800 + 150 \times 20 + 200 \times 40 = 23\,800 \text{ 元}$$

表 6.7

工 序	总时差 $R(i, j)$	工 序	总时差 $R(i, j)$
l	5	p	0
m	5	q	0
n	5	r	0
o	0		

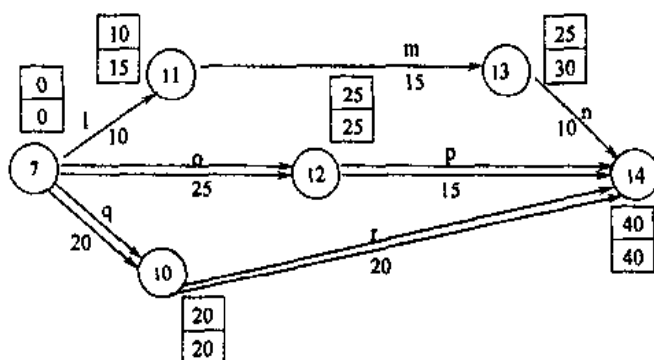


图 6.17

第二步,在关键路线(7)→(12)→(14)上选择 o 工序(7,12)作为压缩工序;在(7)→(10)→(14)上选择 r 工序(10,14)作为压缩工序。

第三步,压缩工序费用率总和为

$$q = [q(7,12) + q(10,14)] = \begin{cases} 210, & \text{当工序(7,12)压缩 5 天内} \\ 250, & \text{当工序(7,12)压缩 6 天以上} \end{cases}$$

无论哪种情况都有 $(q - p) > 0$, 所以不能再压缩工期,任何压缩都将引起总费用上升。最低成本工期为 40 天,相应费用 23 800 元。

6.3.3 工程的时间—资源优化

在使用网络计划技术时,经常遇到工程的时间—资源优化问题,即在资源有限的情况下,如何进行合理调整使资源得以均衡利用,并按期完成工程任务,仅以劳动力资源限制为例,说明如何进行工程的时间—资源优化。

设某项工程的网络计划图由图 6.18 给出。图中箭线上第一个数字为工序时间(单位:周),第二个数字为完成该工序每周所需的劳动力数。通过计算,时间参数也在图 6.18 上给出,该工程工期 $T_E = 9$ 周。

现假设这项工程由一个作业组承包,全组共有 9 名劳力。这个承包组能否在预定的工期 $T_E = 9$ 周内完成工程,各道工序进度应该如何安排才能使劳力资源均衡使用,解决这类问题,一般可按下述四步进行:计算工程每单位时间内所需资源量;作出初始进度横道表及资源需求曲线;进行资源均衡调整求得新的进度计划;评价工程进度计划对资源利用的均衡程度。

对于图 6.18 的工程网络,按上述步骤进行时间—资源优化和调整。

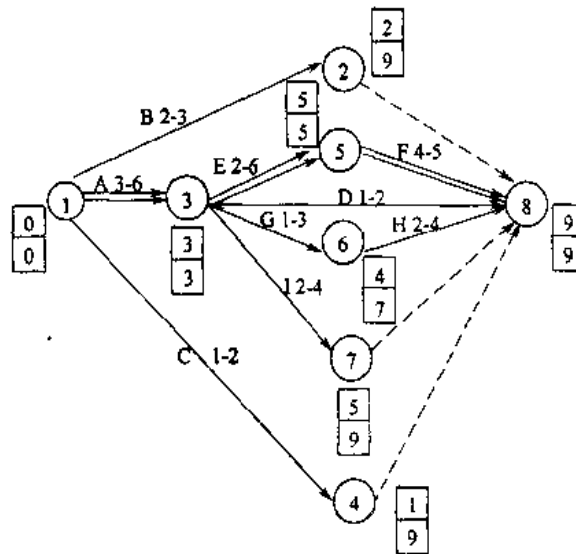


图 6.18

首先,根据数据可算出完成此工程需要周数: $\sum \lambda_i = 79$ 工作周,其中 λ_i 为完成工序所需的工作周数(周 \times 人数)。工程每周所需劳力为

$$\bar{\lambda} = \frac{1}{T_E} \sum \lambda_i = \frac{1}{9} \times 79 = 8.8 \text{ 人}$$

所以适当安排各道工序进程,整个工程有可能在 9 周内完成,相反,要是每周需投入劳力数超过 9 人,则不论怎样调整安排工序进度均不可能在预定工期内完成。

其次,根据网络图及时间参数作出初始进度计划,即把每道工序的开工时间都定在最早可能开工时间 $t_E(i) + 1$,则可以作出各道工序进度的横道表(表 6.8)。表中“—”表示关键工序;“—”表示非关键工序;“...”表示在不影响工期条件下非关键工序允许变动范围。

例如,工序 B 所需的 2 周时间可安排在第一周至第九周这个时间区间内任何持续的 2 周中进行,横线上的数字为工序每周所需劳力数。

表 6.8

工序	1	2	3	4	5	6	7	8	9(周)
A	<u>6</u>	<u>6</u>	<u>6</u>						
B	<u>3</u>	<u>3</u>							
C	<u>2</u>								
D				<u>2</u>					
E				<u>6</u>	<u>6</u>				
F						<u>5</u>	<u>5</u>	<u>5</u>	<u>5</u>
G				<u>3</u>					
H					<u>4</u>	<u>4</u>			
I				<u>4</u>	<u>4</u>				
劳力人数	11	9	6	15	14	9	5	5	5

表 6.8 中最后一行给出了该进度下每周所需劳力资源数。为了更形象描述该进度计划下

资源使用状况,可根据横道表作出劳力资源需求曲线(图 6.19)。资源需求曲线是一个阶梯曲线。其中 t 表示时间; λ 表示资源需求量,即应安排的总人数; λ_{\max} 为资源限制。显然,本例中 $\lambda_{\max}=9$ 人,在第 1、4、5 周中将缺少劳力资源,第 3、7、8、9 周中劳力资源将有剩余。

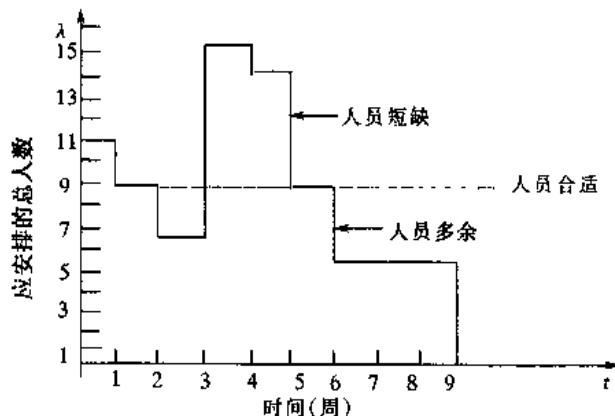


图 6.19

再次,为了均衡使用劳力资源,必须对初始进度计划进行调整,或者更形象地说,必须重新安排工序进度,以使在可能范围把资源需求曲线中凸出部分和凹进部分尽量弄平。对于像本例这样的简单情况,根据非关键工序上的总时差数,可以通过手工计算进行调整而解决,但在工序数目很大时,则必须由电子计算机根据一定程序来进行资源调优。限于篇幅,这里不作一般介绍。在实际工作中,没有必要去追求资源均衡利用的最优状况,对多数实际工程计划来说,只需调整到一个较为满意的状况即可。所以,在实际工作中,往往应用所谓“启发式规划法”解决这一问题。而启发式标准实际上是一种行得通的“约略估计”,通常称约略估计的标准为启发式规划。迄今为止,已经出现许多解决资源平衡网络进度问题的启发式规划法。

启发式方法一般是首先找出那些具有最大总时差的工序,在不拖延整个工程完工期的情况下,试图把这工序尽可能长地往后延缓,例如,首先延缓工序 C 的开工时间至第三周初(工序 C 总时差最大),这样工序 A 和工序 B 可以同时开工,而不超过劳力资源 9 人的限制,继续使用这个方法,就能得到表 6.9 所示的修订好的进度表。

表 6.9

工序	1	2	3	4	5	6	7	8	9(周)
A	6	6	6						
B	3	3							
C			2						
D					2				
E				6	6				
F						5	5	5	5
G				3					
H						4	4		
I								4	4
劳力人数	9	9	8	9	8	9	9	9	9

从表 6.9 上看出,此修订的进度计划在资源利用上有较好的均衡性。

最后,评价一个工程进度计划对资源利用均衡程度一般以单位时间内资源利用量的方差为标准,方差越小,则均衡性越好。若一个工程计划给定后, T_E 为总工期, λ_t 表示第 t 时间单位对资源的需求量, $\bar{\lambda}$ 表示该进度计划下单位时间资源利用量的均值,则根据方差定义,有

$$\lambda_{\sigma^2} = \frac{1}{T_E} \sum_{t=1}^{T_E} (\lambda_t - \bar{\lambda})^2$$

对于图 6.19 所示工程来说,初始进度计划表的 $\lambda_{\sigma^2} = 13.50$ 。这个进度计划下劳力资源利用很不均衡的,在进行时间—资源优化后,修订后进度计划表的 $\lambda_{\sigma^2} = 1.56$ 。这样,它的资源利用就比较均衡了。

6.3.4 工程完工期的概率分析

网络计划中最基本的参数是工序时间。一般来讲,工序时间是一个随机变量,在 PERT 方法中采用三时估计法,即用式子

$$t(i, j) = \frac{a_{ij} + 4m_{ij} + b_{ij}}{6}$$

来计算工序时间,实际就是用这个三时估计值作为工序时间这个随机变量的均值。其方差为

$$\sigma_{ij}^2 = \left(\frac{b_{ij} - a_{ij}}{6} \right)^2$$

一个工程的完工期为关键路线上各工序时间之和。由概率论定理可知,工程的完工期是一个服从正态分布的随机变量。其期望值为关键路线上各工序时间期望之和,即

$$T_E = \sum_{(i,j) \in I} \frac{a_{ij} + 4m_{ij} + b_{ij}}{6}$$

而均方差

$$\sigma = \sqrt{\sum_{(i,j) \in I} \left(\frac{b_{ij} - a_{ij}}{6} \right)^2}$$

其中 I 为一关键路线上工序集合。则工程完工期的分布函数

$$f(T) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^T e^{-\frac{(t-T_E)^2}{2\sigma^2}} dt$$

对于一个网络计划,只要计算出关键路线上的标准差 σ 和完工期的期望值 T_E ,就能对给定某个时间内完成工程的可能性作一概率评价,通过令 $\lambda = \frac{T - T_E}{\sigma}$ 查标准正态分布

$$\Phi(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{t^2}{2}} dt$$

的相应表值,即可知工程在 T 时间内完成的概率为多大。

例如,某工程网络图如图 6.20 所示。先按三时估计法算出每一工序时间 $t(i, j)$ 及方差 σ_{ij}^2 ,并可进一步计算得上述网络之关键路线为

$$\textcircled{1} \xrightarrow{\frac{10-13-15}{12.83}} \textcircled{2} \xrightarrow{\frac{7-8-10}{8.17}} \textcircled{4} \xrightarrow{\frac{8-10-14}{10.33}} \textcircled{6} \xrightarrow{\frac{9-11-13}{11}} \textcircled{7}$$

所得数据列于表 6.10。

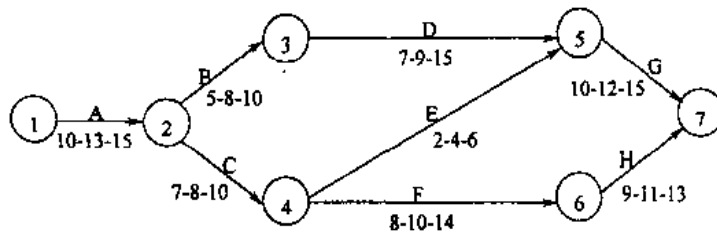


图 6.20

表 6.10

工 序	a	m	b	$t(i, j)$	$\sigma^2(i, j)$
①—②	10	13	15	12.83	0.69
②—④	7	8	10	8.17	0.25
④—⑥	8	10	14	10.33	1.00
⑥—⑦	9	11	13	11	0.44

由表算得,工程完工期的期望值

$$T_E = 12.83 + 8.17 + 10.33 + 11 = 42.33$$

标准差

$$\sigma = \sqrt{0.69 + 0.25 + 1.00 + 0.44} = 1.54$$

若指定要求工程在 40 天内完成,即 $T = 40$,则计算

$$\lambda = \frac{T - T_E}{\sigma} = \frac{40 - 42.33}{1.54} = -1.51$$

查标准正态分布表,可得 $\Phi(-1.51) = 0.066$,则可知工程在 40 天内完成的概率为 0.066。

通过计算工程在各工期内完工的可能性(概率),可比较工程在工期内完工的难易程度。

对于处于极难、困难、较难的完工期,则必须采取一系列的管理措施。当概率为 50% 时,表示通常进行工作也能达到要求,平均完工期与预计完工期一致。而当概率超过 50% 时,则表示这个完工期是容易实现的。

表 6.11

概 率 (%)	工程完工期难易程度评价
0~4	极 难
5~14	困 难
15~49	较 难
50~84	较 易
85~95	容 易
95~100	极 易

另外,如果对工程工期完工的可能性有所要求时,即要求工程完工时间的概率不小于某一数值时,也可根据上述公式算出相应的完工期。上例中若要求完工可能性达 95% 的工期,查

正态分布表 $\Phi(\lambda) = 0.95$,得 $\lambda = 1.6$ 。由 $\lambda = \frac{T - T_E}{\sigma}$,得相应工期为

$$T = T_E + \sigma\lambda = 42.33 + (1.6 \times 1.54) = 45 \text{ 天}$$

习 题

6.1 已知某工程有关资料如表 6.12。

表 6.12

工 序	紧前工序	工序时间	工 序	紧前工序	工序时间	工 序	紧前工序	工序时间
a	—	60	g	b,c	7	m	j,k	5
b	a	14	h	e,f	12	n	i,l	15
c	a	20	i	f	60	o	u	2
d	a	30	j	d,g	10	p	m	7
e	a	21	k	h	25	q	o,p	5
f	a	10	l	j,k	10			

要求:(1)绘制网络图;

(2)计算各项时间参数,求出完工期;

(3)确定关键路线。

6.2 已知如表 6-13 所示资料。

表 6.13

活 动	作业时间	紧前活动	正常完成进度的直接费用(百元)	赶进度一天所需费用(百元)
A	4		20	5
B	8		30	4
C	6	B	15	3
D	3	A	5	2
E	5	A	18	4
F	7	A	40	7
G	4	B,D	10	3
H	3	E,F,G	15	6
合 计			153	
工程的间接费用			5(百元/天)	

要求:(1)按正常情况绘制网络图,求出工期和关键线路;

(2)求出该工程的最低成本工期;

(3)若要求在最低成本工期基础上再提前一天,求出赶工成本最低的赶工方案。

6.3 已知条件如表 6.14,又知间接费率为 15(千元/周),求总费用最低方案及规定工期为 8 周的总费用最低方案。

表 6.14

工作名称	紧前工作	正常工作时间(周)	最短工作时间(周)	直接费率(千元/周)
A	—	2	2	—
B	—	6	4	8
C	A	3	2	12
D	A	5	2	12

续表

工作名称	紧前工作	正常工作时间(周)	最短工作时间(周)	直接费率(千元/周)
E	B、C	4	2	7
F	B、C	3	1	9
G	D、E	3	1	6
H	D、E、F	2	2	—

6.4 已知条件如表 6.15, 其中 a_{ij} 、 m_{ij} 、 b_{ij} 分别是对工作 (i, j) 持续时间所作的乐观、最可能和悲观估计值, 试求该工程在 20 周内完成的概率。

表 6.15

工作名称	紧前工作	a_{ij}	m_{ij}	b_{ij}
A	—	1	2	3
B	—	1	2	3
C	—	1	2	3
D	A	1	2	9
E	A	2	3	10
G	B	3	6	15
H	B	2	5	14
I	D、E	1	4	7
J	C	4	9	20
K	H、J	1	2	9
L	G、I、K	4	4	4

6.5 某项工程有关资料如表 6.16。

表 6.16

工序	紧前工序	平均工序时间(周) $\frac{1}{6}(a+4m+b)$	工序时间方差 $\sigma^2 = \left(\frac{b-a}{6}\right)^2$
A	—	2	1.05
B	—	3	2.25
C	B	2.5	4.34
D	B	6	3.70
E	A、C	20	4.95
F	D、E	4	4.66
G	F	2	4.05

要求: (1) 画出施工网络图, 确定关键工序及期望完工期;

(2) 估计工程在 30 周内完工的概率。

附: 标准正态分布表 $\Phi(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{x^2}{2}} dx$

λ	0.31	0.32	0.33	0.34	0.35
$\Phi(\lambda)$	0.621 7	0.625 5	0.629 3	0.633 1	0.633 8

IV 决策技术

第7章 风险型决策

7.1 决策问题的基本要素及分类

决策这个词起源于英语中的 Decision Making。中文中,“决策”一词应理解为“决定策略”。这里策略是广义的,可以是重大问题的决策,也可以指日常工作、生产中的行动方案。人类在生产、生活中总是不断地遇到问题,决策就是人们为了解决当前或未来可能发生的问题而选择最佳方案的过程。

一般来讲,决策理论是一个比较广义的概念。它涉及到经济理论、组织理论、行为理论等多种学科。本章只介绍决策过程中的数量分析方法,即通常称为“决策分析”的内容。

7.1.1 决策的基本要素

一般来说,任何决策问题都由下面几个基本要素组成:

- ①决策者;
- ②可供选择的方案,用 d_j 表示($j=1,2,\dots,n$);
- ③自然状态,用 θ_i 表示($i=1,2,\dots,m$);
- ④自然状态概率,用 $P(\theta_i)$ 表示($i=1,2,\dots,m$);
- ⑤结局,即各方案在各种可能的自然状态下产生的结果。

当选用 d_j 方案,自然状态为 θ_i 时,产生的结局用 $R(\theta_i, d_j)$ 表示。

在一个决策问题中,至少要有两个以上可供选择的方案。自然状态是指在决策过程中可能出现的情况。自然状态的出现是不以决策者的意志为转移的,是不可控制的。决策分析中结局经常用利润值和损失值表示,如

$$R(\theta_i, d_j) = \begin{cases} u_{ij}, & d_j \text{ 方案在 } \theta_i \text{ 状态下产生的利润值} \\ l_{ij}, & d_j \text{ 方案在 } \theta_i \text{ 状态下产生的损失值} \end{cases}$$

通常结局可用决策损益表表示(表 7.1, 表 7.2)。

表 7.1

利 润 状 态	方 案	d_1	d_2	...	d_n
$P(\theta_1)$	θ_1	u_{11}	u_{12}	...	u_{1n}
$P(\theta_2)$	θ_2	u_{21}	u_{22}	...	u_{2n}
...
$P(\theta_m)$	θ_m	u_{m1}	u_{m2}	...	u_{mn}

表 7.2

损 失 状 态	方 案	d_1	d_2	...	d_n
$P(\theta_1)$	θ_1	l_{11}	l_{12}	...	l_{1n}
$P(\theta_2)$	θ_2	l_{21}	l_{22}	...	l_{2n}
...
$P(\theta_m)$	θ_m	l_{m1}	l_{m2}	...	l_{mn}

下面例子可具体说明决策的各个基本要素。

例 7.1 某唱片磁带公司为了适应市场需求欲扩大生产,计划部门提出三种方案可供公司考虑:①扩大现有工厂;②建立一个新工厂;③将部分产量转包给其他街道工厂生产。对公司来说,最大的不肯定性是未来市场对产品的需求量。经市场预测分析,需求量可能会以如下的可能性出现下面四种状态:①高需求量,出现的可能为 0.2;②中等需求量,出现的可能为 0.5;③低需求量,出现的可能为 0.2;④失败(无需要),出现的可能为 0.1。

根据以往历史资料分析,预计三种方案在各种市场需求状态下,公司能获得的利润值见表 7.3。

表 7.3

结局(收益) (百万元)	方 案	扩大 d_1	新建 d_2	转包 d_3
	自 然 状 态			
$P(\theta_1)=0.2$	高需求 θ_1	500	700	300
$P(\theta_2)=0.5$	中需求 θ_2	250	300	150
$P(\theta_3)=0.2$	低需求 θ_3	-250	-400	-10
$P(\theta_4)=0.1$	无需求 θ_4	-450	-800	-100

问题是在这些数据资料下,公司应选择什么方案最为有利?

上述决策问题中,决策的基本要素如下:

- ①决策者是公司经理;
- ②备选方案是: d_1 为扩大现有工厂, d_2 为建一个新工厂, d_3 为将部分产量转包给其他工厂;
- ③自然状态指市场可能出现状态,即 θ_1 表示需求量高, θ_2 表示需求量中等, θ_3 表示需求量低, θ_4 表示无需求量;

④自然状态概率是指市场各种状态出现的可能性,即 $P(\theta_1)=0.2$, $P(\theta_2)=0.5$, $P(\theta_3)=0.2$, $P(\theta_4)=0.1$ 。

5) 结局是指各方案在各种可能市场状态下的获利值, 由表 7.3 给出。

7.1.2 决策分类

按照决策的自然状态可将决策分为三类。

1. 确定型决策

这类决策问题只可能出现一种确定的自然状态, 每个行动方案在这惟一的自然状态下的结局是可以计算出来的。确定型决策问题是一种逻辑上比较简单的决策, 只需从所有备选方案中, 根据每个方案结局, 选择出一个最好的即可。例如, 第 2 章问题 1 所述线性规划问题就是一个确定型决策问题, 其中:

$X = [x_1 \ x_2]^T$ ——决策方案

$z = 7x_1 + 12x_2$ ——目标函数, 方案 $[x_1 \ x_2]^T$ 下的产值(结局)

$$\begin{cases} 9x_1 + 4x_2 \leq 360 \\ 4x_1 + 5x_2 \leq 200 \\ 3x_1 + 10x_2 \leq 300 \\ x_1, x_2 \geq 0 \end{cases} \quad \text{可行方案必须满足的条件}$$

这是一个确定型决策问题, 可根据各个方案的产值(结局)大小去选择最佳的生产方案。用单纯形法可求得最优解 $x_1 = 20, x_2 = 24$, 最优值为 428。

2. 风险型决策

这类决策问题在决策过程中可以出现多种自然状态 $\theta_i (i = 1, 2, \dots, m)$, 每一个行动方案在不同自然状态下有不同的结局, 且能预先估计出各个自然状态出现的概率 $P(\theta_i) (i = 1, 2, \dots, m)$ 。决策分析的主要讨论对象是风险型决策。

3. 完全不确定型决策

这类决策与风险型决策一样, 在决策过程中可以出现多种自然状态, 但在这类决策问题中, 不能预先估计出各自然状态出现的概率, 所以称之为完全不确定型决策。

7.2 风险型决策

7.2.1 风险型决策的特征

风险型决策有如下特征:

① 决策过程中存在两个以上的自然状态 $\theta_i (i = 1, 2, \dots, m)$, 但究竟会出现哪一种自然状态, 决策者是不能控制的, 但可以事先估计各种自然状态出现的概率 $P(\theta_i) (i = 1, 2, \dots, m)$;

② 存在决策希望达到的明确目标, 例如收益最大或损失最小;

③ 存在多个可供决策者选择的方案;

④ 各个方案在不同自然状态下的结局(收益值或损失值)可以计算出来。

例 7.2 为了开发某种新产品, 某厂需要对生产设备的投资规模作一次决策。设计部门提出三种可供选择的方案: d_1 购买大型设备; d_2 购买中型设备; d_3 购买小型设备。预计新产品投放市场后, 市场对这种产品的需求情况可能有三种, 即有三种可能发生的自然状态: θ_1 需

求量较大; θ_2 需求量中等; θ_3 需求量较小。厂方经销部门根据有关资料与信息预测出市场三种需求情况的出现概率 $P(\theta_1)=0.3, P(\theta_2)=0.4, P(\theta_3)=0.3$ 。厂方的决策目标是使开发新产品的收益最大。厂方估计三种方案 d_1, d_2, d_3 在各种市场状态下的收益用表 7.4 表示。

表 7.4

收益 (万元)		方 案		
		d_1	d_2	d_3
状态及概率				
$P(\theta_1)=0.3$	θ_1	50	30	10
$P(\theta_2)=0.4$	θ_2	20	25	10
$P(\theta_3)=0.3$	θ_3	-20	-10	10

试问在上述情况下厂方应选择何方案最有利?

显然,例 7.2 是典型的风险型决策问题。

7.2.2 期望值准则

决策问题的关键是如何去选择行动方案。一般来讲,首先要对每一个方案作出评价,然后根据这些评价去选择最佳或满意的方案。这种评价在决策分析中是以数量化形式表现的,即对每一个方案作出一个数量化评价值:

$$d_j(\text{方案}) \leftarrow f(d_j)(\text{方案的数量化评价值})$$

于是问题归结为确定方案评价值的准则是什么?也就是决策准则是什么?在风险型决策中,经常采用“期望值准则”。它把一个方案在各种状态下的收益(或损失)的期望值作为这方案的评价值,即有

$$d_j(\text{方案}) \leftarrow f(d_j) = E(d_j) = \sum_{i=1}^m P(\theta_i) u_{ij} (\text{收益期望值})$$

其中 u_{ij} 是 d_j 方案在出现自然状态 θ_i 时的收益值。然后根据各方案评价值——收益或损失期望值的大小去选择方案,即有

$$\max E(d_j) = E^*(d_k), d_k \text{ 为最优方案}$$

上述例 7.2 问题中,先计算三个方案的收益期望值:

$$\begin{aligned} E(d_1) &= P(\theta_1)u_{11} + P(\theta_2)u_{21} + P(\theta_3)u_{31} \\ &= 0.3 \times 50 + 0.4 \times 20 + 0.3 \times (-20) = 17 \end{aligned}$$

$$\begin{aligned} E(d_2) &= P(\theta_1)u_{12} + P(\theta_2)u_{22} + P(\theta_3)u_{32} \\ &= 0.3 \times 30 + 0.4 \times 25 + 0.3 \times (-10) = 16 \end{aligned}$$

$$\begin{aligned} E(d_3) &= P(\theta_1)u_{13} + P(\theta_2)u_{23} + P(\theta_3)u_{33} \\ &= 0.3 \times 10 + 0.4 \times 10 + 0.3 \times 10 = 10 \end{aligned}$$

根据“收益期望值最大准则”得

$$\max E(d_j) = \max\{17, 16, 10\} = 17$$

应选择 d_1 方案,即采用购买大型设备方案最为有利。

例 7.3 某轻工业企业利用剩余生产能力生产一种季节性新产品,自产自销。产品成本每盒 50 元,售出价每盒 80 元;如果当日剩余一盒将损失 20 元。如果当日不能满足市场需要,

除损失销额外,不再有其他损失。可根据去年同期 90 天内类似产品的销售量统计分析未来市场销售需求趋势,大致估计出该产品今年将可能遇到的自然状态(市场需求量) θ_i 及它们出现的概率 $P(\theta_i)$,这由表 7.5 给出。

表 7.5

自然状态 θ_i (日销售量)	100	110	120	130	合 计
去年完成的天数	18	27	36	9	90
估计状态概率 $P(\theta_i)$	$\frac{18}{90}=0.2$	$\frac{27}{90}=0.3$	$\frac{36}{90}=0.4$	$\frac{9}{90}=0.1$	1

该企业拟定今年该产品日产量的备选计划方案为 100 盒、110 盒、120 盒、130 盒,试问该企业应决定日产量为多少最合适?

解 采用最大收益期望准则进行决策。计算各方案在不同的市场销售状态下的收益值,见表 7.6。

表 7.6

收益 (元) / 状态及概率	方 案			
	$d_1=100$ (盒)	$d_2=110$	$d_3=120$	$d_4=130$
$P(\theta_1)=0.2, \theta_1=100$ (盒)	3 000	2 800	2 600	2 400
$P(\theta_2)=0.3, \theta_2=110$	3 000	3 300	3 100	2 900
$P(\theta_3)=0.4, \theta_3=120$	3 000	3 300	3 600	3 400
$P(\theta_4)=0.1, \theta_4=130$	3 000	3 300	3 600	3 900
收益期望 $E(d_j)$	3 000	3 200	3 250	3 100

表 7.6 中收益值 u_{ij} 计算过程如下:

u_{11} 表示日产量 $d_1=100$ 盒,当日市场售出 $\theta_1=100$ 盒时所获的收益

$$u_{11}=30 \times 100=3\,000 \text{ 元}$$

u_{12} 表示日产量 $d_2=110$ 盒,当日市场售出 $\theta_1=100$ 盒,并剩余 10 盒时所获的收益

$$u_{12}=30 \times 100 - 20 \times 10=2\,800 \text{ 元}$$

u_{43} 表示日产量 $d_3=120$ 盒,当日市场需求量为 $\theta_4=130$ 盒时所获收益

$$u_{43}=30 \times 120=3\,600 \text{ 元}$$

根据上述决策收益(表 7.6),计算各生产方案的收益期望值如下:

$$E(d_1=100)=3\,000 \times 0.2 + 3\,000 \times 0.3 + 3\,000 \times 0.4 + 3\,000 \times 0.1=3\,000 \text{ 元}$$

$$E(d_2=110)=2\,800 \times 0.2 + 3\,300 \times 0.3 + 3\,300 \times 0.4 + 3\,300 \times 0.1=3\,200 \text{ 元}$$

$$E(d_3=120)=2\,600 \times 0.2 + 3\,100 \times 0.3 + 3\,600 \times 0.4 + 3\,600 \times 0.1=3\,250 \text{ 元}$$

$$E(d_4=130)=2\,400 \times 0.2 + 2\,900 \times 0.3 + 3\,400 \times 0.4 + 3\,900 \times 0.1=3\,100 \text{ 元}$$

将上述结果填入表 7.6 的最后一行。其中生产方案 $d_3=120$ 盒的收益期望值最大,所以应选择日产量为 120 盒的生产方案。

这个决策问题也可以采用最小损失期望值准则进行决策。这个问题中的损失包括两部分:一部分是由于生产过剩,销售剩余部分在存储过程中变质造成损失;另一部分是因为生产量太小,不能满足市场需求,本应该得到的收益因生产量不足而失去机会造成损失。按照这个

观点,首先计算出每个方案在不同自然状态下(不同销售量下)的损失值,用表 7.7 表示。

表 7.7

损 失 值 \ 方 案	$d_1 = 100(\text{盒})$	$d_2 = 110$	$d_3 = 120$	$d_4 = 130$
状态及概率				
$P(\theta_1) = 0.2, \theta_1 = 100(\text{盒})$	0	200	400	600
$P(\theta_2) = 0.3, \theta_2 = 110$	300	0	200	400
$P(\theta_3) = 0.4, \theta_3 = 120$	600	300	0	200
$P(\theta_4) = 0.1, \theta_4 = 130$	900	600	300	0
$E(d_j)$	420	220	170	320

表 7.7 中损失值 l_{ij} 计算如下:

日产量方案为 $d_1 = 100$ 盒、日销售量为 $\theta_1 = 100$ 盒时的损失值 $l_{11} = 0$ 。

日产量方案为 $d_3 = 120$ 盒、日销售量为 $\theta_1 = 100$ 盒时的损失值 $l_{13} = 20 \times (120 - 100) = 400$ 元(报废性损失)。

日产量方案为 $d_1 = 100$ 盒、日销售量为 $\theta_4 = 130$ 盒时的损失值 $l_{41} = 30 \times (130 - 100) = 900$ 元(机会损失)。

根据决策损失值表(表 7.7),计算各生产方案的期望损失值 $E(d_j)$,并填在表 7.7 最后一行内,即

$$E(d_1) = 0 \times 0.2 + 300 \times 0.3 + 600 \times 0.4 + 900 \times 0.1 = 420 \text{ 元}$$

$$E(d_2) = 200 \times 0.2 + 0 \times 0.3 + 300 \times 0.4 + 600 \times 0.1 = 220 \text{ 元}$$

$$E(d_3) = 400 \times 0.2 + 200 \times 0.3 + 0 \times 0.4 + 300 \times 0.1 = 170 \text{ 元}$$

$$E(d_4) = 600 \times 0.2 + 400 \times 0.3 + 200 \times 0.4 + 0 \times 0.1 = 320 \text{ 元}$$

其中方案 $d_3 = 120$ 盒的损失期望值 $E(d_3) = 170$ 元,为最少。按损失期望值最小准则,应选择日产量 120 盒的方案。这个结论与按最大收益期望准则决策所得的结论是一致的。事实上,还可以用期望值公式验证这一一般性结论,有兴趣的读者不妨自证。

必须指出,在使用期望值准则决策时,期望值并不代表必然能实现的数值。因此,在一次性决策中按期望值准则选择的方案,也不一定就是效果最好的方案。从统计学角度来看,以期望值作为评选方案优劣的数量标准还是合理的,即如果这类决策问题重复出现多次,则收益期望值大的方案优于收益期望值小的方案。

7.2.3 决策树

风险型决策可以借助于一种名为“决策树”的图形做出。它不仅可以处理单阶段决策问题,而且可以有效地解决一些多阶段决策问题。

1. 决策树的结构

决策树是一种由结点和分支构成的由左向右横向展开的树状图形。

(1) 结点

决策树中的结点可分为下列三种。

1) 决策结点 通常用方块□表示。由决策结点引出若干分支,每个分支表示一个方案,称

方案分支。

2) 状态结点 用圆形○表示。由状态结点引出若干分支,每个分支表示一个自然状态,称为状态分支或概率分支。

3) 结局结点 通常用三角形△表示。它表示一个方案在一个自然状态下的结局。

(3) 分支

决策树中的边称为分支,分支分为两类。

1) 方案分支 由决策结点引出的分支,每一分枝表示一个方案。

2) 概率分支 由状态结点引出的分支,每个分支表示一种自然状态。每个概率分支上标明它表示的自然状态及自然状态概率。

图 7.1 表示一简单的决策树。

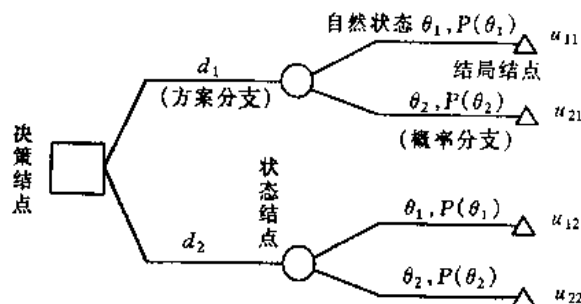


图 7.1

2. 利用决策树进行决策的步骤

① 绘制决策树。

② 自右至左计算各个方案的期望值,将计算结果标在方案分支右端状态结点旁。

③ 根据各方案期望值大小进行选择,将在收益期望值小的方案分支上画上删除号,表示应删去。这个过程称为“剪支”,所保留下来的分支即为最优方案。

3. 实例

(1) 单阶段决策问题

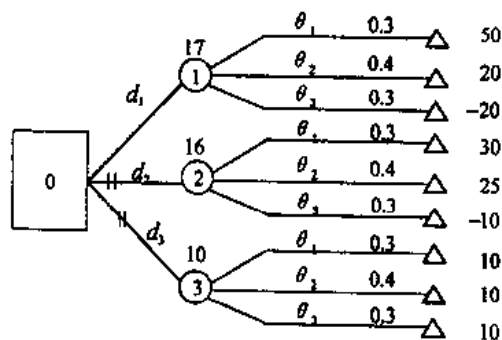


图 7.2

以例 7.2 为例,画出决策树进行决策,如图 7.2 所示。

方案分支 d_1 的期望值即右端状态结点 1 的期望值,记为

$$E(1) = 0.3 \times 50 + 0.4 \times 20 + 0.3 \times (-20) = 17,$$

同理可得

$$E(2) = 0.3 \times 30 + 0.4 \times 25 + 0.3 \times (-10) = 16,$$

$$E(3) = 0.3 \times 10 + 0.4 \times 10 + 0.3 \times 10 = 10$$

根据期望值准则,将 d_2 和 d_3 分支画上删除号表示剪去,剩余 d_1 分支即为最优方案。

(2)多阶段决策问题

例 7.4 有一个化工原料厂,由于某项工艺不够好,产品成本较高,现在计划将该项工艺加以改进。取得新工艺有两条途径:一是自行研究,估计成功的可能性是 0.6;二是从国外引进,估计谈判成功的可能性是 0.8。不论研究成功还是谈判成功,生产规模都考虑两种方案,一是产量不变,一是增加产量。如果自行研究和谈判都失败,则仍采用原工艺进行生产,并保持原产量不变。

根据市场预测,估计今后 5 年内这种产品跌价的可能性是 0.1,保持中等价的可能性是 0.5,涨价的可能性是 0.4,各状态下的收益值见表 7.8,试用决策树法进行决策。

表 7.8

益 损 值 价 格 状 态	方 案	按原工艺生产	引进技术成功(0.8)		自行研究成功(0.6)	
			产量不变	增加产量	产量不变	增加产量
价格低落(0.1)		-100	-200	-300	-200	-300
价格中等(0.5)		0	50	50	0	-250
价格高涨(0.4)		100	150	250	200	600

解 ①从左往右画出决策树。决策点 1 引出两个方案分支:引进技术和自行研究。两个方案分支的端点是状态结点 2 和 3。从结点 2 引出两个状态分支:失败(概率 0.2)或成功(概率 0.8),其中失败分支的端点是状态结点 4,而成功分支的端点是决策结点 5。由于引进技术失败后维持原工艺不变,不需要作决策了,因此,只需从结点 4 引出价格低、价格中、价格高的三个状态分支,并标上相应概率,然后在右端结局结点上标上相应的收益值。对于成功分支右端的决策结点 5,又需引出两个方案分支:产量不变的方案分支和增加产量的方案分支。产量不变的方案分支的右端结点 8 引出三个状态分支,填上相应概率并在各状态分支右端的结局结点上标上相应收益值。增加产量的方案分支的右端结点 9,同样引出三个状态分支,也填上各自状态概率,并在尾部结点上标上相应的收益值。

对于状态结点 3,也引出两个状态分支,即自行研究成功和自行研究失败,分别标上状态概率 0.6 和 0.4。其中成功分支的端点是决策结点 6。它引出产量不变和增加产量两个方案分支,每个方案分支右端结点又各自有三个状态分支。失败分支右端是状态结点 7,它同状态结点 4 一样,标上三种状态的概率和相应的收益值。

从图 7.3 看出,例 7.4 问题中的决策有两级。第一级要作产量不变和增加产量两个方案的选择和决策,第二级作出引进技术和自行研究这两个方案间的选择和决策。

例 7.5 某公司为建一工厂制订了两个方案:一个方案是一次完成建设大厂;另一个方案是分两步,先建个小厂,二年后再考虑是否进行扩建。假设资源条件约束无特殊要求,公司有关部门已有如下资料。

①销售部门提供了市场需求预测(表 7.9)。

表 7.9

10 年预测		概 率
前 2 年需求量	后 8 年需求量	
高	高	0.6
高	低	0.1
低	低	0.3
低	高	0

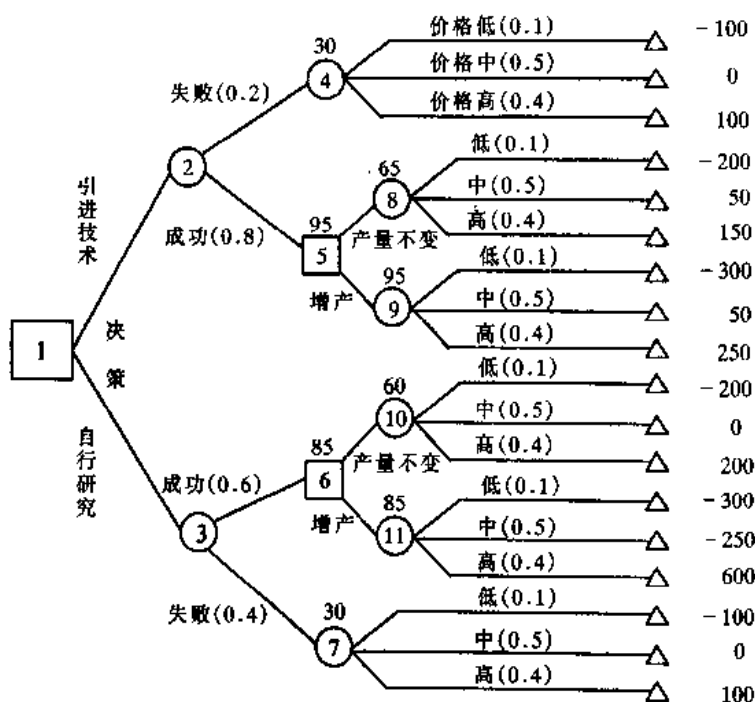


图 7.3

②基建、生产、财务部门提供成本效益预测为建大厂投资 300 万元,建小厂投资 130 万元,建小厂再扩充规模投资 220 万元。三种建厂方案在市场需求量高和市场需求量低的情况下,每年的收益值见表 7.10。

表 7.10

收 益		方 案		
状 态		建大厂	建小厂	扩建小厂
	需求量高	100 万	40 万	70 万
	需求量低	10 万	30 万	5 万

公司该怎样决策才能使 10 年内总收益最大?

解 这个问题是一个二阶段决策问题。第一阶段需确定是建大厂还是建小厂;若第一阶段选择建小厂时,二年后需决定是否要扩建小厂规模。使用决策树方法来处理这个问题。

①根据资料计算各方案下的收益值和未来各种市场状态的出现概率。

首先,计算收益值如表 7.11 所示。

表 7.11

10 年收益		方 案		
市 场 状 态		建大厂 (投资 300 万)	先建小厂(投资 130 万)	
			二年后扩建(投资 220 万)	二年后不扩建
前 2 年高需求后 8 年高需求(0.6)		$100 \times 10 = 1\,000$ 万	$40 \times 2 + 70 \times 8 = 640$ 万	$40 \times 10 = 400$ 万
前 2 年高需求后 8 年低需求(0.1)		$100 \times 2 + 10 \times 8 = 280$ 万	$40 \times 2 + 5 \times 8 = 120$ 万	$40 \times 2 + 30 \times 8 = 320$ 万
前 2 年低需求后 8 年低需求(0.3)		$10 \times 10 = 100$ 万	$30 \times 2 + 5 \times 8 = 100$ 万	$30 \times 10 = 300$ 万

然后,计算 10 年期内各种市场状态出现的概率如下:

$$P(\text{前2年高需求}) = P(\text{前2年高需求且后8年高需求}) + P(\text{前2年高需求且后8年低需求}) = 0.6 + 0.1 = 0.7$$

同理,由事件间的互斥性得: $P(\text{前2年低需求}) = 0.3$ 。再由条件概率公式可得:

$$P(\text{后8年高需求}|\text{前2年高需求}) = \frac{0.6}{0.7} = 0.86$$

$$P(\text{后8年低需求}|\text{前2年高需求}) = 1 - 0.86 = 0.14$$

$$P(\text{后8年低需求}|\text{前2年低需求}) = \frac{0.3}{0.3} = 1$$

②绘制决策树(图7.4)。

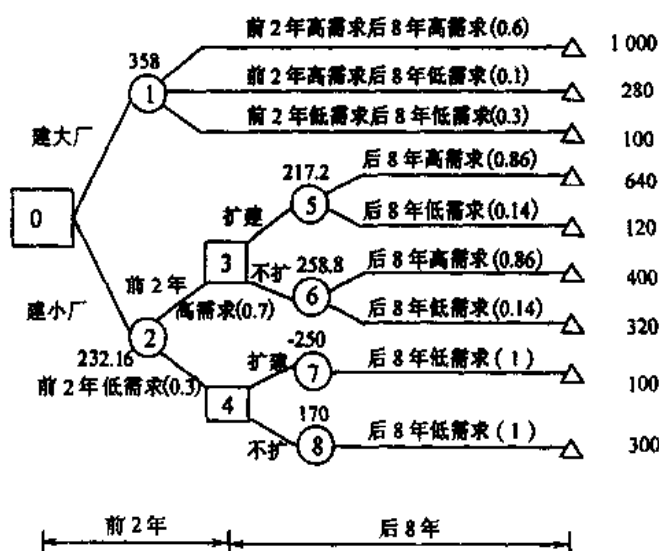


图 7.4

③计算各方案分支(右端状态结点)的收益期望值,并在各决策结点对方案分支进行剪支。

$$E(5) = 640 \times 0.86 + 120 \times 0.14 - (220 + 130) = 217.2 \text{ 万}$$

$$E(6) = 400 \times 0.86 + 320 \times 0.14 - 130 = 258.8 \text{ 万}$$

在决策结点3处应把扩建分支剪去。

计算结点7和结点8的收益期望值。得

$$E(7) = 100 - (220 + 130) = -250 \text{ 万}$$

$$E(8) = 300 - 130 = 170 \text{ 万}$$

在决策结点4处应将扩建分支剪去。

计算状态结点1的收益期望值,得

$$E(1) = 1000 \times 0.6 + 280 \times 0.1 + 100 \times 0.3 - 300 = 358 \text{ 万}$$

计算结点2的收益期望值,得

$$E(2) = 258.8 \times 0.7 + 170 \times 0.3 = 232.16 \text{ 万}$$

经比较后,在决策结点0处将建小厂分支剪去,即应选择建大厂方案最为有利。

决策树分析法本质上还是一种期望值准则决策方法。对于一些比较复杂的决策问题,使用决策树比较有效,对于多级决策问题尤为方便。决策树方法现已得到广泛的使用,它有几个明显的优点:

①决策树反映整个决策过程,使决策者能够以一定顺序,有步骤、有条理地去周密考察各有关因素;

②决策树方法要求决策者考虑检验所有可能的结果,合意和不合意的都同样要进行考虑;

③决策树以一种非常直观、简明的方式,用图形把决策过程形象地显示出来,使更多的人了解这个决策过程,以使用集体智慧进行决策;

④决策树方法可以通过计算机进行计算和模拟,可以观察分析这些备选方案的变化对最终结果产生的影响。

7.2.4 贝叶斯决策

在风险型决策中,对自然状态出现概率 $P(\theta_i)$ 估计的正确程度直接影响到决策中收益期望值。为了更好地进行决策,在条件许可的情况下,往往需要进一步补充新信息。补充信息可以通过进一步调查、试验、咨询得到,而为了获得这些补充信息需支付一定费用。获得新信息后,可根据这些补充信息修正原先对自然状态出现概率的估计值 $P(\theta_i)$,并利用修正的概率分布重新进行决策。由于这种概率修正主要根据概率论中的贝叶斯(Bayes)定理进行,故称这种决策为贝叶斯决策。贝叶斯决策通常可以分为三步进行。

1)先验分析 决策者首先根据资料及经验对各自然状态出现的概率 $P(\theta_i)$ 作出估计,称为先验概率,然后依据先验概率分布及期望值准则作出决策,选择出最优方案,并得出相应最优期望值,记为 EMV^* (先)。

2)预验分析 在补充新信息前,先对补充信息是否合算作出分析,从而决定是否补充新信息。

3)后验分析 根据获得的新信息,对先验概率分布进行修正,得到后验概率分布,在此基础上作出决策,并计算出补充信息的价值。

1. 先验分析

首先估计自然状态 θ_i 出现概率为 $P(\theta_i)$ ($i=1,2,\dots,m$), u_{ij} 为方案 d_j 在 θ_i 状态下的收益值,见表 7.12。

表 7.12

收 益 值 状态及概率	方 案				
		d_1	d_2	...	d_n
$P(\theta_1) \theta_1$		u_{11}	u_{12}	...	u_{1n}
$P(\theta_2) \theta_2$		u_{21}	u_{22}	...	u_{2n}
...
$P(\theta_m) \theta_m$		u_{m1}	u_{m2}	...	u_{mn}

根据期望值准则,计算各方案收益期望值:

$$E(d_j) = \sum_{i=1}^m P(\theta_i) u_{ij} \quad (j=1,2,\dots,n)$$

相应最优决策方案及最优期望值

$$\max_j E(d_j) = E(d_k) = EMV^* \text{ (先)}$$

2. 预验分析

信息的价值在于它能提高决策的最大期望收益值,但是如果为获得信息所花费的费用超过它所能提高的期望收益值,这种补充信息是不合算的。所有信息中最好、最理想的信息自然是完全可靠、准确的信息,即这种信息预报某自然状态出现,则在实际中必定出现这自然状态,这种信息称为完全信息。在预验分析中,首先估算出完全信息的价值(任何信息的价值均不会超过完全信息的价值),并以它为一个标准。如果补充信息费用 C_i 远远小于完全信息之价值,则可认为这种补充信息是合算的。反之,如果信息费用 C_i 接近甚至超过完全信息的价值,则可认为这种补充信息是不合算的。

当完全信息预报出现 θ_k 状态时,问题变为确定型决策问题。最优方案显然应由下式确定:

$$\max_j \{u_{kj}\}$$

在完全信息下,决策所能获得的最大收益期望值

$$EPPI = \sum_{k=1}^m P(\theta_k) \cdot \max_j \{u_{kj}\}$$

显然, $EPPI$ 与 EMV^* (先) 之间的差额就是得到了完全信息而使期望收益增加的部分。这个值就是该问题的完全信息价值,简记为 $EVPI$ 。

$$EVPI = EPPI - EMV^* (\text{先})$$

需要指出,十分完善的信息是很难得到的,有时甚至根本无法获得完全信息。因此,算出的完全信息价值,常常只是作为支付信息费用的一个上限,也是决定是否有必要进一步获取情报信息的依据。

例 7.6 计算例 7.2 所述问题中完全信息的价值。

解 由例 7.2 所解,已算出先验分析中最大期望收益值为 17 万元,即有

$$EMV^* (\text{先}) = \max E(d_j) = 17 \text{ 万元}$$

在完全信息下:

预报 θ_1 出现,由 $\max\{50, 30, 10\} = 50$,选 d_1 为最优方案;

预报 θ_2 出现,由 $\max\{20, 25, 10\} = 25$,选 d_2 为最优方案;

预报 θ_3 出现,由 $\max\{-20, -10, 10\} = 10$,选 d_3 为最优方案。

完全信息下最大期望收益值为

$$EPPI = 0.3 \times 50 + 0.4 \times 25 + 0.3 \times 10 = 28 \text{ 万元}$$

完全信息的价值

$$EVPI = EPPI - EMV^* (\text{先}) = 28 - 17 = 11 \text{ 万元}$$

3. 后验分析

后验分析工作由补充新信息、计算修正概率、重新决策、计算补充信息的价值四部分组成。

(1) 补充新信息

补充新信息一般是通过对 x_1, x_2, \dots, x_s 共 s 个状态的调查、试验,预报其中哪一个将出现,同时通过资料获取条件概率 $P(x_j | \theta_i)$,即实际出现自然状态 θ_i 而预报 x_j 的概率。

(2) 修正概率

在已知先验概率 $P(\theta_j) (j=1, 2, \dots, m)$ 及条件概率 $P(x_j | \theta_i) (j=1, 2, \dots, s; i=1, 2, \dots, m)$ 的基础上,利用贝叶斯公式可计算出修正概率,即后验概率

$$P(\theta_j | x_i) = \frac{P(\theta_j)P(x_i | \theta_j)}{\sum_{j=1}^m P(\theta_j)P(x_i | \theta_j)}$$

上述后验概率的计算也可用表格进行(表 7.13)。

表 7.13

I 状态概率 $P(\theta_j)$	II 条件概率 $P(x_i \theta_j)$					III $P(x_i \cap \theta_j)$					IV 后验概率 $P(\theta_i x_j)$				
	x_1	...	x_i	...	x_s	x_1	...	x_i	...	x_s	x_1	...	x_i	...	x_s
$\theta_1 \quad P(\theta_1)$	$\cdots P(x_i/\theta_1) \cdots$					$\cdots P(\theta_1)P(x_i/\theta_1) \cdots$					$\cdots P(\theta_1/x_i) \cdots$				
$\theta_2 \quad P(\theta_2)$	$\cdots P(x_i/\theta_2) \cdots$					$\cdots P(\theta_2)P(x_i/\theta_2) \cdots$					$\cdots P(\theta_2/x_i) \cdots$				
\vdots	\vdots					\vdots					\vdots				
$\theta_m \quad P(\theta_m)$	$\cdots P(x_i/\theta_m) \cdots$					$\cdots P(\theta_m)P(x_i/\theta_m) \cdots$					$\cdots P(\theta_m/x_i) \cdots$				
$P(x_1) \cdots P(x_i) \cdots P(x_s)$															

表 7.13 中后验概率计算过程如下:

(a)表中 I 为自然状态概率, II 部分为条件概率 $P(x_i | \theta_j)$, 可以依次填入表内;

(b)表中 III 部分为概率 $P(x_i \cap \theta_j)$, III 中第 i 列各元素可由 I 和 II 中第 i 列相应元素相乘而得, 例如 $P(x_i \cap \theta_2) = P(\theta_2) \cdot P(x_i | \theta_2)$;

(c)表中 III 部分最后一行元素 $P(x_i)$ 是 III 中第 i 列各元素之和, 即

$$P(x_i) = \sum_{j=1}^m P(\theta_j) \cdot P(x_i | \theta_j)$$

(d)表中 IV 部分为后验概率, IV 中第 i 列元素是由 III 中第 i 列各元素分别除以 $P(x_i)$ 而得, 即为

$$P(\theta_j | x_i) = \frac{P(\theta_j) \cdot P(x_i | \theta_j)}{P(x_i)} \quad (j=1, 2, \dots, m)$$

(3)重新决策

根据已得的后验概率, 可预先作出决策的框架。假设补充信息预报将出现 x_k 状态, 则使用后验修正概率分布 $P(\theta_j | x_k)$ ($j=1, 2, \dots, m$), 计算各方案的期望收益值, 并依期望值准则进行决策, 得

$$E(d_j | x_k) = \sum_{i=1}^m P(\theta_i | x_k) u_{ij} \quad (j=1, 2, \dots, n)$$

$$\max E(d_j | x_k) = E(d_{j^*} | x_k)$$

选择 d_{j^*} 为预报 x_k 时的最优方案, 相应最大期望收益值记为

$$E(x_k) = E(d_{j^*} | x_k)$$

一旦得到补充信息预报, 即可按上述方式进行决策。

(4)计算补充信息价值

根据已计算出的补充信息预报各状态出现的概率 $P(x_i)$ ($i=1, 2, \dots, s$), 可计算出后验分析中最大期望收益值

$$EMV^*(后) = \sum_{i=1}^s P(x_i) E(x_i)$$

显然, 获得补充信息后, 期望收益值增加了 $EMV^*(后) - EMV^*(先)$ 。这样, 补充信息的

价值为 $EMV^*(后) - EMV^*(先)$ 。由此,可将补充信息的价值与获得信息所付出的代价进行对比,从而作出正确决策。

应当指出的是,补充信息通常具有不确定性,因而,这样的信息是不完全的,或说不是绝对准确的,这种信息也称为抽样信息。因此,与完全信息相比,补充信息价值不会大于完全信息价值。

例 7.7 某工程项目按合同应在三个月内完工,其施工费用与工程完工期有关。假定天气是影响工程能否按期完工的决定因素,如果天气好,工程能按时完工,施工单位可获利 5 万元;如果天气不好,不能按时完工,施工单位将被罚款 1 万元;若不施工就要损失窝工费 2 千元。根据过去的经验,在计划施工期内天气好的可能性为 30%。为了更好地掌握天气情况,可请气象中心作进一步天气预报,并提供同一时期天气预报资料,这需支付信息资料费 0.08 万元。从提供的资料中可知,气象中心对好天气预报准确性为 80%,对坏天气预报准确性为 90%。问该如何进行决策?

解 采用贝叶斯决策方法。

I. 先验分析

根据已有资料作出决策损益表(表 7.14)

表 7.14

收益值(万) 状 态	方 案	d_1	d_2
		施 工	不施工
好天气 θ_1 (0.3)		5	-0.2
坏天气 θ_2 (0.7)		-1	0.2
$E(d_i)$		0.8	-0.2

根据期望值准则选择方案 d_1 施工有利,相应最大期望收益值 $EMV^*(先) = 0.8$ 万元。

II. 预验分析

计算完全信息下最大期望收益值 $EPPI$ 和完全信息的价值 $EVPI$,得

$$EPPI = 0.3 \times 5 + 0.7 \times (-0.2) = 1.36 \text{ 万元}$$

$$EVPI = 1.36 - 0.8 = 0.56 \text{ 万元}$$

而信息资料费 $C_s = 0.08 \leq 0.56$,所以初步认为请气象中心提供信息和资料是合算的。

III. 后验分析

a) 补充信息 气象中心将提供预报此时期内两种天气状态 x_1 (好天气)、 x_2 (坏天气)将会出现哪一种状态(本例中补充信息预报的状态正好和原自然状态一致)。

从气象中心提供的同期天气资料可得知条件概率:

天气好且预报天气也好的概率 $P(x_1 | \theta_1) = 0.8$;

天气好而预报天气不好的概率 $P(x_2 | \theta_1) = 0.2$;

天气坏而预报天气好的概率 $P(x_1 | \theta_2) = 0.1$;

天气坏且预报天气也坏的概率 $P(x_2 | \theta_2) = 0.9$ 。

b) 计算后验概率分布 根据全概率公式和贝叶斯公式,可计算下列后验概率。预报天气好与天气坏的概率分别为

$$P(x_1) = P(\theta_1) \cdot P(x_1 | \theta_1) + P(\theta_2) \cdot P(x_1 | \theta_2)$$

$$=0.3 \times 0.8 + 0.7 \times 0.1 = 0.31$$

$$P(x_2) = P(\theta_1) \cdot P(x_2 | \theta_1) + P(\theta_2) \cdot P(x_2 | \theta_2)$$

$$=0.3 \times 0.2 + 0.7 \times 0.9 = 0.69$$

预报天气好且天气实际也好的概率

$$P(\theta_1 | x_1) = \frac{P(\theta_1) \cdot P(x_1 | \theta_1)}{P(x_1)} = 0.3 \times 0.8 / 0.31 = 0.77$$

预报天气好而天气实际不好的概率

$$P(\theta_2 | x_1) = \frac{P(\theta_2) \cdot P(x_1 | \theta_2)}{P(x_1)} = 0.7 \times 0.1 / 0.31 = 0.23$$

预报天气不好而实际天气好的概率

$$P(\theta_1 | x_2) = \frac{P(\theta_1) \cdot P(x_2 | \theta_1)}{P(x_2)} = 0.3 \times 0.2 / 0.69 = 0.09$$

预报天气不好且实际天气也不好的概率

$$P(\theta_2 | x_2) = \frac{P(\theta_2) \cdot P(x_2 | \theta_2)}{P(x_2)} = 0.7 \times 0.9 / 0.69 = 0.91$$

上述计算也可用表格形式进行,见表 7.15。

表 7.15

先验概率 $P(\theta_j)$		条件概率(x_i/θ_j)		$P(x_i \cap \theta_j)$		后验概率 $P(\theta_j/x_i)$	
		x_1	x_2	x_1	x_2	x_1	x_2
θ_1	0.3	0.8	0.2	0.24	0.06	0.77	0.09
θ_2	0.7	0.1	0.9	0.07	0.63	0.23	0.91

$$P(x_1) = 0.31 \quad P(x_2) = 0.69$$

c) 后验决策 若气象中心预报天气好(x_1),则每个方案的最大期望收益值

$$E(d_1 | x_1) = 0.77 \times 5 + 0.23 \times (-1) = 3.62$$

$$E(d_2 | x_1) = 0.77 \times (-0.2) + 0.23 \times (-0.2) = -0.2$$

选择 d_1 即施工的方案,相应于预报 x_1 时的最大期望收益值

$$E(x_1) = 3.62$$

若气象中心预报天气不好(x_2)时,各方案的最大期望收益值

$$E(d_1 | x_2) = 0.09 \times 5 + 0.91 \times (-1) = -0.46$$

$$E(d_2 | x_2) = -0.2$$

选择 d_2 即不施工方案,相应于预报 x_2 时的最大期望收益值

$$E(x_2) = -0.2$$

d) 计算补充信息的价值 在有气象中心的补充信息及资料条件下,后验决策的最大期望收益值

$$EMV^*(后) = P(x_1) \cdot E(x_1) + P(x_2) \cdot E(x_2)$$

$$= 0.31 \times 3.62 + 0.69 \times (-0.2) = 0.9842$$

气象中心提供补充信息的价值

$$EMV^*(后) - EMV^*(先) = 0.9842 - 0.80 = 0.1842 \text{ 万元}$$

通过计算知,花费了信息费 $C_z = 0.08$ 万元,提高了决策期望收益 0.1842 万元,这种花费

是值得的,这也验证了预验分析中的判断。

整个决策过程可以用决策树(图 7.5)表示。

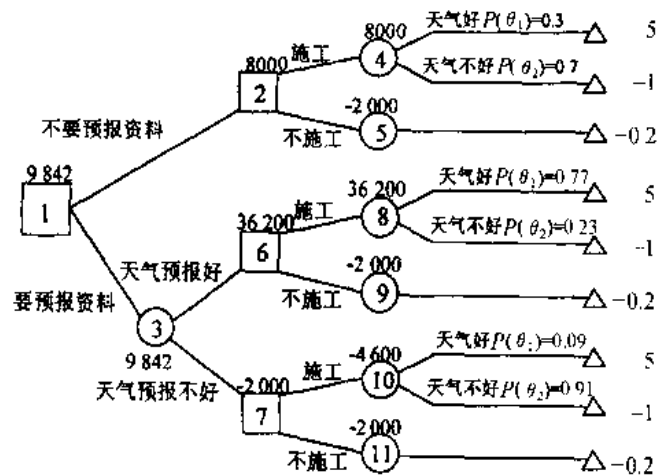


图 7.5

例 7.8 某钻井队面临两种决策:钻(d_1)和不钻(d_2)。若钻井会遇到两种情况:有油(θ_1)和无油(θ_2)。钻井出油能获利 4 万元,反之不出油要亏损 2 万元。据技术部门预测该地区出油概率为 0.3。为了获得更多信息,可以请地质部门勘探此地区的地质结构。这种勘探结果可能会出现三种结果: x_1 (有闭合结构)、 x_2 (可能有闭合结构)、 x_3 (无闭合结构)。地质部门同时能提供以往的勘探资料(表 7.16),勘探和资料费用 0.2 万元。试问该钻井队是否进行勘探,并如何选择决策方案。

表 7.16

实际出现状态 \ 勘探结果	$P(x_i \theta_j)$		
	x_1	x_2	x_3
θ_1	0.4	0.4	0.2
θ_2	0.1	0.5	0.4

解 先验分析,计算各方案期望收益值

$$E(d_1) = 0.3 \times 4 + 0.7 \times (-2) = -0.2 \quad E(d_2) = 0$$

则应选择 d_2 , 不进行钻井, 相应 EMV^* (先) = 0。

预验分析, 计算完全信息的价值, 得

$$EPPI = 0.3 \times 4.0 + 0.7 \times 0 = 1.2 \quad EVPI = 1.2 - 0 = 1.2 \text{ 万元}$$

信息费用 $C_s = 0.2 \ll 1.2$, 则请地质部门勘探来补充信息是合算的。

后验分析, 先计算后验概率, 见表 7.17。

表 7.17

$P(\theta_j)$	$P(x_i / \theta_j)$			$P(x_i \cap \theta_j)$			$P(\theta_j / x_i)$		
	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
θ_1 0.3	0.4	0.4	0.2	0.12	0.12	0.06	0.63	0.26	0.18
θ_2 0.7	0.1	0.5	0.4	0.07	0.35	0.28	0.37	0.74	0.82

$$P(x_1) = 0.19 \quad P(x_2) = 0.47 \quad P(x_3) = 0.34$$

然后进行决策。当勘探报告 x_1 出现时,计算各方案期望值

$$E(d_1|x_1) = 4.0 \times 0.63 + (-2) \times 0.37 = 1.78$$

$$E(d_2|x_1) = 0$$

则选择方案 d_1 (钻井), 相应 $E(x_1) = 1.78$

当勘探报告 x_2 出现时, 计算

$$E(d_1|x_2) = 4 \times 0.26 - 2.0 \times 0.74 = 0.44$$

$$E(d_2|x_2) = 0$$

则选择 d_2 方案(不钻), 相应 $E(x_2) = 0$ 。

当勘探报告 x_3 出现时, 计算

$$E(d_1|x_3) = 4 \times 0.18 + (-2) \times 0.82 = -0.92$$

$$E(d_2|x_3) = 0$$

则选择 d_2 方案(不钻), 相应 $E(x_3) = 0$ 。

后验最大期望收益值

$$EMV^*(后) = 0.19 \times 1.78 + 0.47 \times 0 + 0.34 \times 0 = 0.34 \text{ 万元}$$

补充信息的价值

$$EMV^*(后) - EMV^*(先) = 0.34 - 0 = 0.34 \text{ 万元}$$

通过计算知, 花费了信息费 $C_i = 0.2$ 万元, 提高了决策期望收益 0.34 万元, 这种花费是划算的。

7.2.5 效用值准则

1. 期望值准则的缺陷

在本节前面所讲的风险型决策问题的分析方法中, 是以损益期望值准则为选取策略的标准, 且认为相等的损益期望值对各个决策者的吸引力都一样。换言之, 是它的价值看成了对谁都一样的客观价值。但在大多数情况下, 这样做是有欠缺的, 单纯依据期望值准则进行决策难以达到决策者希望的效果。一方面, 由于在实际风险型决策问题中, 决策者的选择是受他对决策风险的态度所影响的, 并不完全依据于损益期望值的大小。对于同样一件非确定性的事情来说, 不同的人, 因为他们的社会经济地位不同, 价值观念各异, 对待风险的态度会有较大差别, 选用的策略也会不同。所以, 在决策过程中, 完全排斥决策者主观因素的影响和固守纯数量化的决策准则是不科学的; 另一方面, 最大期望收益和最小期望损失, 这两种准则适用于一次决策多次重复的情况, 因为它们是平均意义下的最大收益和最小损失。对于实际决策过程中经常只做一次或少数几次试验的情况, 它就不是十分恰当了。最后, 在某些场合下, 有些策略的效用通常很难用数量指标表示, 或者即使能用数量指标表示, 却无法进行相互比较, 此时就不能用期望值准则。

下述事例可以从一个侧面说明期望值准则的这种欠缺。

例 7.9 某稀有金属矿为了扩大生产量, 制定了两个增产方案: 一是新建一矿井; 一是对原有老矿进行技术改造。根据现有的技术力量和设备, 估计对老矿井进行技术改造成功的把握较大, 即采用改旧方案, 5 年后能稳获 250 万元利润。而对新建矿增产把握不大, 如果成功, 产量能大幅度增长, 预计 5 年内能获利 2 000 万元, 若失败则要亏损 1 000 万元。根据以往经

验,开发成功和失败的概率各为 0.5。试问该采用哪个方案?这个问题的决策损益表见表 7.18。

表 7.18

收益(万) 状态及概率		方 案	
		d_1 新 建	d_2 改 旧
成功 θ_1	0.5	2 000	250
失败 θ_2	0.5	-1 000	250

若根据收益期望值大小作决策准则,计算

$$E(d_1) = 2\,000 \times 0.5 + (-1\,000) \times 0.5 = 500 \text{ 万}$$

$$E(d_2) = 250 \text{ 万}$$

则该采用 d_1 新建矿方案为好,但是,这是一个一次性利害关系重大的决策,有的决策者会觉得,虽然新建方案成功能获利 2 000 万元,但却要冒很大风险,一旦失败则要损失 1 000 万元,所以,他宁愿采用改旧方案 d_2 稳拿 250 万收益,而不愿承担损失的风险。在这些决策者心目中,稳拿 250 万元的满意程度要超过风险下的期望收益 500 万的满意程度。这就说明在实际决策中,决策者并不是机械地按期望收益值大小来选择方案,而是要依赖决策者对风险的态度。一般来讲,当同一决策问题重复多次进行决策时,或决策所冒的风险较小时,决策者对方案的选择与按期望值大小选择是一致的,而当决策只进行一次或少数几次时,或包含有较大风险时,则对方案的选择与期望损益值大小两者之间就会有相当的距离。

鉴于期望值准则存在这样或那样的缺陷,有必要提出一种新的决策准则——效用值准则。以下首先阐明效用与效用函数的概念。

2. 效用与效用函数

为了定量地描述决策者对风险的偏好和厌恶的程度,可使用“效用”这样一个概念衡量同一期望值在不同人主观上的价值。效用值是风险下损益值在决策者心目中的满意程度的衡量尺度。实质上,效用值大小是反映决策者对风险态度的数量指标,习惯上最大效用值用 1 表示,最小效用值用 0 表示。在一个决策过程中,把各种结局中最大收益值的效用定为 1(最满意),把最小收益值的效用定为 0(最不满意),并可通过与决策者对话去确定决策者对每一个损益值的效用值,即能确定一个决策者的效用函数 $U(x)$,并有

$$0 \leq U(x) \leq 1$$

决策者的效用函数在以横坐标表示损益值、以纵坐标表示效用值的直角坐标系中为一条曲线。它描述了特定的损益值(财富)水平和与之相对应的满足程度之间的关系,称为决策者的效用曲线。通过效用函数或效用曲线可确定任一损益值的效用值。效用函数一定程度上反映决策者对风险的态度、决策者偏向及评价标准。

确定效用函数(曲线)的基本方法有两种:一种是直接提问法;另一种是对比提问法,又称心理试验法。

1) 直接提问法 向决策者提出一系列的问题,要求决策者进行主观权衡并作出回答。如向某决策者提问:“今年你收入 2 万元,你是满意的,那么收入增至多少,你会加倍满意?”若该决策者回答 5 万元。这样反复提问和回答,便可绘出该决策者的收入效用曲线。由于这种提问与回答是十分含糊的,很难做到确切,所以实际中应用较少。

2)心理试验法 设效用函数的自变量 x 代表决策过程中的损益值,函数值 $U(x)$ 是 x 在决策者心目中的效用值。进行心理试验法时,首先确定甲、乙两个方案,然后询问决策者选择哪一个方案。选择甲方案,表示他可以概率 p 得到 a 元,或以 $(1-p)$ 的概率损失 b 元,不妨用符号 $(p, a; 1-p, b)$ 表示甲方案。若选择乙方案,则可无任何风险稳得 c 元,且 $a > c > b$,不妨用符号 c 表示乙方案。

若在某条件下,决策者认为甲、乙两方案等价时,记为: $c \sim (p, a; (1-p), b)$ 。用等式可表示为:

$$pU(a) + (1-p)U(b) = U(c)$$

确切地讲,此时该决策者认为 c 的效用值等价于 a 、 b 的效用期望值。

从上述等式可见,等式中共有 a 、 b 、 c 、 p 四个变量。若其中任意三个已知时,向决策者提问第四个变量应取何值? 并请决策者作出主观判断第四个变量应取的值是多少。如此反复回答,便可据此绘制出该决策者的效用曲线。

一般而言,对比提问的方式大致有以下三种:

- ①每次固定 a 、 b 、 c 的值,改变 p ,问决策者“ p 取何值时,认为甲方案与乙方案等价”;
- ②每次固定 p 、 c 、 a (或 b) 的值,改变 b (或 a) 的值,问决策者“ b (或 a) 取何值时,认为甲方案与乙方案等价”;
- ③每次固定 p 、 a 、 b 的值,改变 c 的值,问决策者“ c 取何值时,认为甲方案与乙方案等价”。

在上述三种提问方式中,第三种运用较多。具体是:取 $p = \frac{1}{2}$,固定 a 、 b ,利用 $\frac{1}{2}U(a) + \frac{1}{2}U(b) = U(c)$,改变 c 值。若决策者认为甲方案的期望效用值高于乙方案的效用值(左端大于右端),则适当提高 c 值,继续询问。反之,则适当降低 c 值,继续问答。直至乙方案调整到某一 c 值时,上述等式成立,即在决策者心目中稳拿当前的 c 元之效用值与选择甲方案的效用值一样。此时,在坐标系中给出一个点 $(c, U(c))$ 。然后,再通过固定新的 a 、 b 值继续这一过程,在坐标系中给出其他点(三点即可,连同最大及最小效用值点,共五点),最后利用光滑曲线将这些点联结起来,就得到该决策者的效用曲线。

以下详细说明心理试验法的具体步骤:

- ①确定决策过程中最大收益值 c_0 和最小收益值 c_1 ,选择概率 $p = \frac{1}{2}$,则有 c_0 和 c_1 的效用值分别为

$$U(c_0) = 1 \text{ 和 } U(c_1) = 0$$

- ②与决策者对话,用试问法确定 c_2 ,满足

$$c_2 \sim (\frac{1}{2}, c_0; \frac{1}{2}, c_1)$$

则 c_2 的效用值与 $(\frac{1}{2}, c_0; \frac{1}{2}, c_1)$ 的效用值相等,即有

$$U(c_2) = \frac{1}{2}U(c_0) + \frac{1}{2}U(c_1) = \frac{1}{2} \times 1 + \frac{1}{2} \times 0 = \frac{1}{2}$$

- ③继续用试问法确定 c_3 ,满足

$$c_3 \sim (\frac{1}{2}, c_0; \frac{1}{2}, c_2)$$

同理可得

$$U(c_3) = \frac{1}{2} U(c_0) + \frac{1}{2} U(c_2) = \frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}$$

④继续用试问法确定 c_4 , 满足

$$c_4 \sim (\frac{1}{2}, c_1; \frac{1}{2}, c_2)$$

同理可得

$$U(c_4) = \frac{1}{2} U(c_1) + \frac{1}{2} U(c_2) = \frac{1}{2} \times 0 + \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

⑤通过与决策者对话确定了 5 个益损值的效用值, 即得效用曲线上 5 个点 $(c_0, 1)$ 、 $(c_1, 0)$ 、 $(c_2, \frac{1}{2})$ 、 $(c_3, \frac{3}{4})$ 、 $(c_4, \frac{1}{4})$ 。在直角坐标上标出这 5 个点, 并用光滑曲线连起来便成一条效用曲线, 见图 7.6。

如对例 7.9 描述的决策问题, 使用心理试验法去确定某决策者的效用曲线, 可得

① $c_0 = 2\,000$, $c_1 = -1\,000$, 有

$$U(2\,000) = 1, U(-1\,000) = 0$$

②通过询问确定 $c_2 = 0$, 即有

$$0 \sim (\frac{1}{2}, 2\,000; \frac{1}{2}, -1\,000)$$

可确定 $U(0) = \frac{1}{2}$ 。

③通过询问确定 $c_3 = 800$, 即有

$$c_3 \sim (\frac{1}{2}, 2\,000; \frac{1}{2}, 0)$$

可确定 $U(800) = \frac{3}{4}$ 。

④通过询问确定 $c_4 = -600$, 即有

$$c_4 \sim (\frac{1}{2}, -1\,000; \frac{1}{2}, 0)$$

可确定 $U(-600) = \frac{1}{4}$ 。

⑤把上述 5 点用光滑曲线连起来, 就构成这个决策者的效用曲线(图 7.7)。

决策者的效用曲线可分为三种类型, 如图 7.8 所示。不同类型的效用曲线反映决策者对风险的不同态度。

I 型曲线所反映决策者的特点是: 他认为稳得一个收益值的效用要大于有风险的相同甚至更大的期望收益值的效用, 也就是他宁愿选择稳得 500 万元的方案, 而不愿选择具有期望值

$$0.5 \times 2\,000 + 0.5 \times (-1\,000) = 500 \text{ 万}$$

的方案 $(\frac{1}{2}, 2\,000; \frac{1}{2}, -1\,000)$ 。这种类型决策者对收益反应较迟缓, 而对损失则比较敏感。

这是一种规避风险、谨慎从事、不求大利的保守类型的决策者。如在保险业, 投保人为了避免

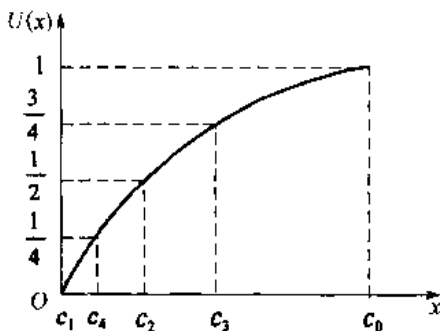


图 7.6

风险,在保险公司愿意承担全部损失的情况下,投保人愿意付出的净保费多于平均的期望损失。由图 7.8 知,风险规避者的效用函数建立在两个假设上:一是收益数量的增加导致满足程度的上升,即有 $U'(x) > 0$;二是收益增加时边际效用降低,即有 $U''(x) < 0$ 。

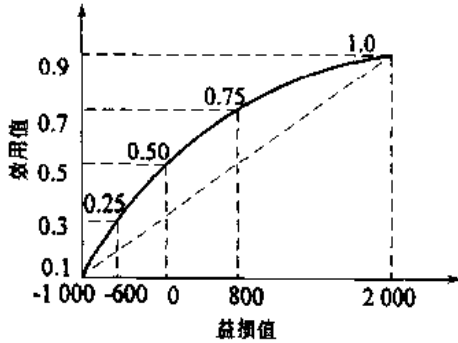


图 7.7

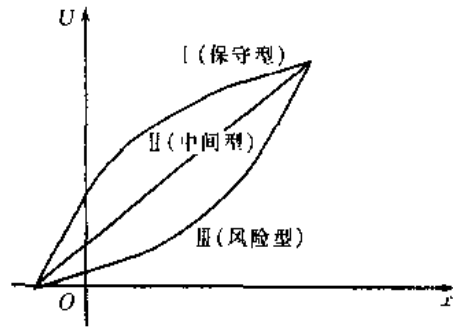


图 7.8

Ⅲ型曲线反映的决策者的特点正好相反,他宁愿选择带有风险下期望收益为 $0.5 \times 2\,000 + 0.5 \times (-1\,000) = 500$ 万的

方案 $(\frac{1}{2}, 2\,000; \frac{1}{2}, -1\,000)$, 而不愿意选择稳得 500 万的方案。这种决策者的注意力着重放在收益上,是一种不怕风险、谋求大利、乐于进取的人。同保守型决策者的效用曲线相对照,读者不难归纳出这类决策者的效用函数所具有的特点。

Ⅱ型曲线反映的是一种中间型决策者。这类决策者严格按照期望收益准则选择方案。其效用曲线是线性的,称这类决策者是风险中性的。

虽然不同的人对待风险的态度不同,从而,他们可能主要属于以上三种决策类型之一。但

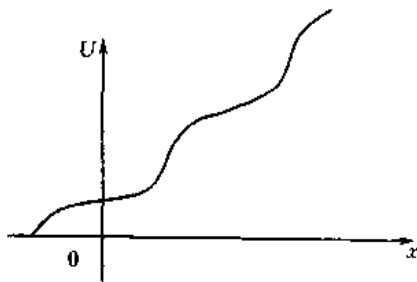


图 7.9

对某一个人来说,在不同的时间和条件下,他对风险的态度有可能发生变化。例如,有的人在起初对较小的收益不太有兴趣,但随着收益的进一步增加,吸引力就会逐步增大,从而引起他对风险的态度发生变化。可是当达到某一目标后,他的要求得到了满足,就可能变得不愿承担风险了。然而,当收益的继续增加使他可望达到一个更高的目标时,他又有可能不顾冒更大的风险去争取,如此等等。因此,上述三种情况只是三种典型,某一决策者可能兼有三种类型。图 7.9 示出这种

情形。

3. 效用值准则

利用决策者的效用函数进行决策时,依据的准则称为效用准则。

例 7.10 某公司在建厂问题上有两种方案 d_1 (建大厂) 和 d_2 (建小厂), 建大厂需投资 300 万元, 建小厂需投资 160 万元。两种方案的生产期都定为 10 年, 估计在此期间, 产品销路好的可能性为 0.7。两方案的年度损益表见表 7.19。试问应该选择哪一方案为宜?

表 7.19

收益 (万元)		方 案	d_1 建大厂	d_2 建小厂
状态及概率				
销路好 θ_1	0.7		100	40
销路差 θ_2	0.3		-20	10

解 使用效用值准则进行决策,步骤如下。

①根据数据可以计算各方案在 10 年内不同状态下的收益值(表 7.20)。

表 7.20

收益 (万元)		方 案	d_1 建大厂(投资 300 万)	d_2 建小厂(投资 160 万)
状态及概率				
销路好 θ_1	0.7		$100 \times 10 - 300 = 700$	$40 \times 10 - 160 = 240$
销路差 θ_2	0.3		$(-20) \times 10 - 300 = -500$	$10 \times 10 - 160 = -60$

②作出决策树(图 7.10)。

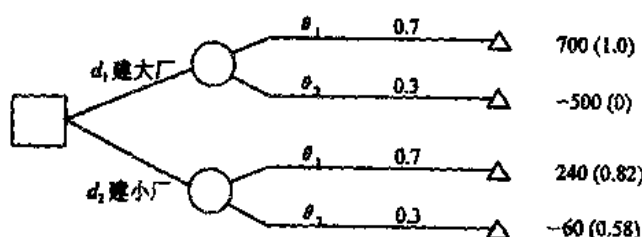


图 7.10

③通过和决策者的对话,作出决策者的效用曲线(图 7.11)。

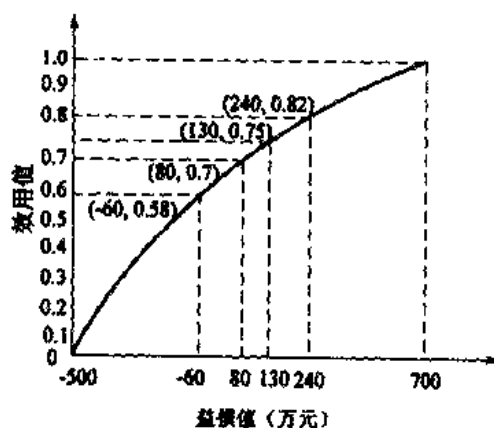


图 7.11

从确立的效用曲线上找出各收益值的效用值为: $U(700) = 1$, $U(-500) = 0$, $U(240) = 0.82$, $U(-60) = 0.58$ 。把这些效用值填在决策树各损益值边括弧内。

④计算各方案的期望效用值。建大厂方案的期望效用值 $= 0.7 \times 1.0 + 0.3 \times 0 = 0.7$, 建小厂方案的期望效用值 $= 0.7 \times 0.82 + 0.3 \times 0.58 = 0.75$ 。根据效用值准则,应选效用期望值大的方案 d_2 , 即选择建小厂方案。

7.3 完全不确定情况下的决策

在风险型决策问题中,虽然预先不知道会发生哪种自然状态,但是可以通过历史资料或主观估计方案的方法得到每种自然状态发生的概率。在许多实际决策问题中,对自然状态的信息掌握很少,不足以估算出自然状态的分布概率,这种决策称为完全不确定型决策。在这种情况下,只知道几种自然状态可能发生,但并不知道这些状态发生的概率,因而不能运用风险型决策模型,这时很大程度上要依靠决策者的决策偏好进行决策。

下面介绍几种常用的完全不确定型决策方法。

7.3.1 悲观法——小中取大原则

这是用保守观点分析和决定问题,决策者对客观情况总是抱悲观态度,总觉得不会万事如意。所以为了保险起见,总是把事情的结果估计得很不利。决策者认为冒失的行动容易造成重大失误,最好还是从最不利的情况出发,向最好的方向努力,力求做到有备无患。他是在各种最坏的结果中找出一个好一点的方案,因此又叫小中取大决策法。根据这种思想,决策者用一个方案在各种自然状态下最小的收益值作为这个方案的评价值,即有

$$\text{方案 } d_j \leftarrow \text{评价值 } f(d_j) = \min_i u_{ij}$$

然后根据每个方案的评价值去选择最优方案,即根据

$$\max_j f(d_j) = \max_j \min_i u_{ij}$$

选择评价值最大的方案为最优方案。

例 7.11 现有三个行动方案 d_1 、 d_2 、 d_3 ,三个自然状态 θ_1 、 θ_2 、 θ_3 ,不知道自然状态发生的概率,相应决策收益表由表 7.21 给出。试用悲观法求最优方案。

表 7.21

收 益 状 态	方 案			
		d_1	d_2	d_3
θ_1		50	30	10
θ_2		20	25	10
θ_3		-20	-10	10
评价值 $f(d_j)$		-20	-10	(10)

解 悲观法基本步骤如下。

①计算每个方案的评价值 $f(d_j)$,即为每个方案在各自然状态下的最小收益值:

$$f(d_1) = \min\{50, 20, -20\} = -20$$

$$f(d_2) = \min\{30, 25, -10\} = -10$$

$$f(d_3) = \min\{10, 10, 10\} = 10$$

②根据方案评价值选择最优方案,即从各最小收益值中求最大值:

$$\max\{f(d_1), f(d_2), f(d_3)\} = \max\{-20, -10, 10\} = 10$$

则应选择方案 d_3 为最优方案。

若讨论的问题是机会损失时,悲观法则是先取各方案中最大的机会损失值,再在其中选最小值确定为最优方案。

7.3.2 乐观法——大中取大原则

这种方法的基本思想是对客观情况总抱乐观态度。一些决策者认为,即使出现不利情况也未必会有多大损失,而一旦出现最有利的情况却能得到很大的收益,由此产生大中取大的原则。决策者用一个方案在各种自然状态下最大的收益值作为这个方案的评价值,即有

$$\text{方案 } d_j \leftarrow \text{评价值 } f(d_j) = \max_i u_{ij}$$

然后选最大评价值的方案为最优方案,即根据

$$\max_j f(d_j) = \max_j \max_i u_{ij}$$

选择最优方案。

例 7.12 用乐观法求解例 7.11。

解 乐观法基本步骤如下。

①计算各方案评价值 $f(d_j)$,即计算各方案在各自然状态下最大收益值

$$f(d_1) = \max\{50, 20, -20\} = 50$$

$$f(d_2) = \max\{30, 25, -10\} = 30$$

$$f(d_3) = \max\{10, 10, 10\} = 10$$

②根据各方案评价值,选择评价值最大的方案为最优方案,即有

$$\max\{f(d_1), f(d_2), f(d_3)\} = \max\{50, 30, 10\} = 50$$

则应选择方案 d_1 为最优方案。其决策过程见表 7.22。

表 7.22

收 益 状 态	方 案			
		d_1	d_2	d_3
θ_1		50	30	10
θ_2		20	25	10
θ_3		-20	-10	10
评价值 $f(d_j)$		(50)	30	10

7.3.3 乐观系数法

这种方法的基本想法是认为悲观法太保守而乐观法太冒险,主张折衷考虑,确定一个数字表示乐观程度,称为乐观系数,用 α 表示,并规定 $0 \leq \alpha \leq 1$ 。决策者用一个方案在各自然状态下最大收益值乘以 α 加上最小收益值乘以 $(1-\alpha)$ 的和,作为该方案的评价值,即有

$$\text{方案 } d_j \leftarrow \text{评价值 } f(d_j) = \alpha \cdot \max_i u_{ij} + (1-\alpha) \min_i u_{ij}$$

然后再选择评价值大的方案为最优方案,即根据

$$\max_j \{f(d_j)\} = \max_j \{ \alpha \cdot \max_i u_{ij} + (1-\alpha) \min_i u_{ij} \}$$

选择最优方案。

例 7.13 用乐观系数法求解例 7.11。

解 选择乐观系数为 $\alpha=0.4$, 计算各方案之评价值

$$f(d_1)=0.4 \times 50 + 0.6 \times (-20) = 8$$

$$f(d_2)=0.4 \times 30 + 0.6 \times (-10) = 6$$

$$f(d_3)=0.4 \times 10 + 0.6 \times 10 = 10$$

根据

$$\max\{f(d_1), f(d_2), f(d_3)\} = \max\{8, 6, 10\} = 10$$

则应选择方案 d_3 为最优方案, 整个决策过程由表 7.23 所示。

表 7.23

收 益 状 态	方 案			
		d_1	d_2	d_3
θ_1		50	30	10
θ_2		20	25	10
θ_3		-20	-10	10
评价值 $f(d_j)$		$0.4 \times 50 + 0.6 \times (-20) = 8$	$0.4 \times 30 + 0.6 \times (-10) = 6$	$0.4 \times 10 + 0.6 \times 10 = 10$

7.3.4 最小后悔值法——大中取小原则

当某一种自然状态出现时, 决策者必然选择收益值最大的方案。如果决策者事先未选取这一方案, 而选择了其他方案, 事后就会后悔。某一自然状态下, 最大收益值与其他收益值之差描绘了后悔程度, 称为后悔值。这个方法要在决策时考虑减少将来的后悔。它的计算方法首先是在各种自然状态下, 以最大收益值减去其他收益值得到这自然状态下各方案的后悔值。在 θ_i 状态下, 选择 d_j 方案的后悔值为

$$r_{ij} = \max_j u_{ij} - u_{ij}$$

所有后悔值构成后悔值矩阵 $[r_{ij}]_{m \times n}$ 。

在后悔值方法中, 决策者用一个方案在各自然状态下最大的后悔值作为这个方案的评价值, 即有

$$\text{方案 } d_j \leftarrow f(d_j) = \max_i r_{ij}$$

然后选择评价值最小的方案为最优方案, 即根据

$$\min_j f(d_j) = \min_j \max_i r_{ij}$$

选择最优方案, 所以也称为最小后悔值方法。

例 7.14 用最小后悔值法求解例 7.11。

解 最小后悔值方法步骤为:

① 计算出后悔矩阵 $(r_{ij})_{m \times n}$;

② 计算每一方案评价值 $f(d_j) = \max_i r_{ij}$;

③ 选择最小评价值 $\min_j f(d_j)$, 确定最优方案。

决策过程可由表 7.24 所示。

表 7.24

方 案 状 态	收益值表			后悔值表		
	d_1	d_2	d_3	d_1	d_2	d_3
θ_1	50*	30	10	0	20	40
θ_2	20	25*	10	5	0	15
θ_3	-20	-10	10*	30	20	0
	评价值 $f(d_j)$			30	(20)	40

7.3.5 等可能性方法——平均原则

这种方法的基本思想是:在一些决策问题中,决策者既不能估计出各自然状态出现的概率,又不能认为某一自然状态比其他状态出现可能性大,就只好认为每种情况出现的可能性一样,如有 n 种自然状态出现,则认为每种自然状态出现概率为 $p(\theta_i) = \frac{1}{n}$ ($i=1,2,\dots,n$)。这样就把完全不确定型决策问题化为风险型决策问题解决。

对于一个完全不确定型问题,使用不同的方法,可能会得出不同的最优方案。实际决策问题是很复杂的,究竟应选用什么方法,应视具体问题而定。比如对待灾难性事件,应考虑最不利情况,使遭到损失最小。在其他情况下应有不同考虑,通常可根据具体情况选用几种不同的方法,然后将所得结果进行分析和比较,从而作出最后的选择。

习 题

7.1 某地方书店希望订购最新出版的好图书。根据以往经验,新书的销售量可能为 50、100、150 或 200 本。假定每本新书的订购价为 4 元,销售价为 6 元,剩书的处理价为每本 2 元。要求:

- (1)建立损益矩阵;
- (2)分别用悲观法、乐观法及等可能法决定该书店应订购的新书数字;
- (3)建立后悔矩阵,并用后悔值法决定书店应订购的新书数。

7.2 上题中如书店据以往统计资料预计新书销售量的规律见表 7.25。(1)分别用期望值法和后悔值法决定订购数量;(2)如某市场调查部门能帮助书店调查销售量的确切数字,该书店愿意付出多大的调查费用?

表 7.25

需求数	50	100	150	200
占的比例	20%	40%	30%	10%

7.3 题 7.1 中,如果该地方书店货币(M)的效用函数

$$U(M) = \sqrt{\frac{M+1\,000}{1\,000}}$$

- (1)建立效用值表;
- (2)利用题 7.2 中给出的各种需求量的比例数字重新决定该书店应当订购新书的最优数字。

7.4 某邮局要求当天收寄的包裹当天处理完毕。根据以往记录统计,每天收寄包裹的情况见表 7.26。

表 7.26

收寄包裹数	41 ~ 50	51 ~ 60	61 ~ 70	71 ~ 80	81 ~ 90
占的比例	10%	15%	30%	25%	20%

已知每个邮局职工平均每小时处理 4 个包裹,每小时工资为 5 元。规定每人每天实际工作 7 小时。如加班工作,每小时工资额增加 50%,但加班时间每人每天不得超过 5 小时(加班以小时计,不足 1 小时的以 1 小时计算)。试用期望值法确定该邮局最优雇佣工人的数量。

7.5 在一台机器上加工制造一批零件共 10 000 个,如加工完后逐个进行修整,则全部可以合格,但需修整费 300 元。如不进行修整据以往资料统计,次品率情况见表 7.27。一旦装配中发现次品时,需返工修理费为每个零件 0.50 元。要求:

(1)分别用期望值和后悔值法决定这批零件要不要修整;

(2)为了获得这批零件中次品率的正确资料,在刚加工完的一批 10 000 件中随机抽取 130 个样品,发现其中有 9 件次品,试修正先验概率,并重新按期望值和后悔值法决定这批零件要不要修整。

表 7.27

次品率(θ)	0.02	0.04	0.06	0.08	0.10
概率 $P(\theta)$	0.20	0.40	0.25	0.10	0.05

7.6 有一种游戏分两阶段进行。第一阶段,参加者需先付 10 元,然后从含 45% 白球和 55% 红球的罐子中任摸一球,并决定是否继续第二阶段。如继续需再付 10 元,根据第一阶段摸到的球的颜色在相同颜色罐子中再摸一球。已知白色罐子中含 70% 蓝球和 30% 绿球,红色罐子中含 10% 的蓝球和 90% 的绿球。当第二阶段摸到为蓝色球时,参加者可得奖 50 元,如摸到的是绿球或不参加第二阶段游戏的均无所得。试用决策树法确定参加者的最优策略。

7.7 某公司有 50 000 元多余资金,如用于某项开发事业估计成功率为 96%,成功时一年可获利 12%,但一旦失败,有丧失全部资金的危险。如把资金存放到银行中,则可稳得年利 6%。为获取更多情报,该公司求助于咨询服务,咨询费用为 500 元,但咨询意见只是提供参考,帮助下决心。据过去咨询公司类似 200 例咨询意见实施结果,情况见表 7.28。试用决策树法分析:

(1)该公司是否值得求助于咨询服务;

(2)该公司多余资金应如何合理使用?

表 7.28

咨询意见 \ 实施结果	投资成功	投资失败	合 计
	投资成功	投资失败	合 计
可以投资	154 次	2 次	156 次
不宜投资	38 次	6 次	44 次
合 计	192 次	8 次	200 次

7.8 某工厂正在考虑是现在还是明年扩大生产规模问题。由于可能出现的市场需求情况不一样,预期利润也不同。已知市场需求为高(E_1)、中(E_2)、低(E_3)的概率及不同方案时的预期利润(单位:万元)如表 7.29 所示。对该厂来说损失 1 万元效用值为 0,获利 10 万元效用值为 100,对以下事件效用值无差别:①肯定得 8 万元或 0.9 概率得 10 万和 0.1 概率失去 1 万;②肯定得 6 万元或 0.8 概率得 10 万和 0.2 概率失去 1 万;③肯定得 1 万元或 0.25 概率得 10 万和 0.75 概率失去 1 万。

表 7.29

方 案	事 件	E_1	E_2	E_3
	概 率	$P(E_1)=0.2$	$P(E_2)=0.5$	$P(E_3)=0.3$
现在扩大		10	8	-1
明年扩大		8	6	1

要求:(1)建立效用值表;

(2)分别根据实际盈利额和效用值按期望值法确定最优决策。

第8章 库存决策

8.1 库存决策的基本概念

库存决策又称存贮论,是管理技术中的一个重要分支。在生产活动及日常生活中经常发生供应和需求之间不协调的情况。物资储存(即库存)就是解决和协调供应和需求之间矛盾的一种手段。如工厂为了避免由于原料供应不足而停工,就必须储备一定数量的原料;商店为了避免商品脱销也必须储备一定量的商品。无论是营利单位,或者是非营利单位,物资的储存是极普遍的现象。例如:水库储备一定量的水,以备旱季使用;银行储备一定量现金,保证提款的需要;血库储备一定量血液,为抢救病人备用;军队储备一定量军需品及弹药,以备一旦敌人发动战争能予以反击;家庭储备一定量粮食和食品,以备日常生活所需。

然而,库存物资是否越多越好?工厂生产过程中原料储存越多,停工待料现象就越少,从而减少停工引起的损失。但是,原料储存过多相应储存费用也增加,同时又占用了大量的流动资金,造成经济损失。一般来讲,库存量不足会造成缺货损失,而库存量过大又会造成物质积压,库存费用增大,流动资金占用过大。因此,如何制定一个合理的库存控制方针,使缺货风险和库存过多之间得以平衡,是存贮论要研究和解决的主要问题。库存控制方针首先就是要解决:库存物资每次补充多少,库存物资隔多少时间补充一次?

为了解决这两个问题,一般是对所研究的实际储存系统建立一个相应的模型,由此模型的分析研究得出最佳的库存方针。评价库存方针的优劣,通常是以与这库存方针相应的成本费用大小作标准。对一个企业来讲,实际库存工作是由两方面组成,其一是日常物资储存的收、发、存等具体工作,账目的计算和记录,各种数据和档案的保管工作可以由电子计算机进行,其二是确定合理的订购时间和订购数量。存贮论的研究主要是针对第二方面工作的。

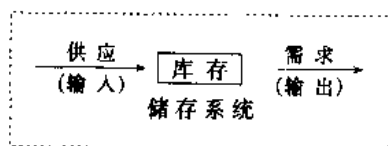


图 8.1

图 8.1 表示一般储存系统模型。在讨论具体模型前,先引入一些有关储存系统的基本概念。

1. 需求

需求是指对某种储存物资的需要,用单位时间内对这种物资的需求量进行描述,需求是储存系统的输出。需求

可以是均匀连续的,也可以是间断的;需求量可以是确定的,也可以是随机的。

2. 补充

存储量由于需求输出而不断减少,必须加以补充。补充就是储存系统的输入。补充的方式有订货或生产。补充可以是瞬时进行,也可以是均匀进行。

1) 批量 补充(输入)往往采用以一定数量为一批的方式进行,每一批补充数量为批量。

2) 补充间隔 两次补充之间的时间。

3) 提前期(拖后期) 从提出订货到收到订货之间的时间间隔。提前期可以是确定的,也可以是随机的。

3. 储存系统的费用

1) 储存费用 这包含仓库保管费用, 占用流动资金利息, 储存物资的变质损失等等, 以单位储存物资在单位时间内所需的费用计算。

2) 准备费用 这是指每一次订货或每组织一次生产所必须的固定费用, 例如每次订货的手续费、出差费等, 每次生产的准备、结束费等。它与订货数量或生产数量无关。

3) 货物成本费用(生产可变成本费用) 它是指货物本身的价格, 或者是与生产产品数量有关的可变成本费用。

4) 缺货损失费用 因库存量不足, 发生供不应求而造成的利润损失。例如商品缺货失去销售机会的损失、停工待料的损失以及不能履行合同的罚款。

4. 储存策略

决定多少时间补充一次以及每次补充多少的策略。通常有三种类型。

1) t 循环策略 每隔 t 时间补充存储量 Q 。

2) (s, S) 策略 每当存储量下降至 s 时, 即刻补充, 使存储量达 S 。

3) (t, s, S) 混合策略 每隔 t 时间检查库存量 x 。当 $x > s$ 不补充; 当 $x \leq s$ 时进行补充, 使库存量达到 S 。

为了确定一个储存系统的最佳储存策略, 首先对实际储存系统建立相应的数学模型。这模型既要不太复杂又要能反映出实际系统的主要本质特点。长期以来人们已建立了许多有效的储存模型。这些模型根据数据(例如需求量、提前期等)是确定的还是随机的, 可分成两大类: 一类为确定型储存模型; 另一类是随机型储存模型。

8.2 确定型库存模型

本节讨论的三个库存模型中的参数, 例如需求量、提前期等都是确定的。这类模型是将复杂的实际情况简化了, 所以便于分析。在这些模型讨论中, 评价一个储存策略优劣的数量指标是这个储存策略产生的总费用。所谓最佳储存策略就是使总费用最小的策略。

8.2.1 经济批量 EOQ 库存模型(模型 1)

这个模型假设:

① 需求量是均匀、连续的, 单位时间内需求量是常数 R , 则 t 时间内需求量是 Rt ;

② 提前期为零, 补充是瞬时的, 即一订货即刻就能得到补充, 每次补充量 Q 是不变的, 这意味着, 需要时马上就可以补充, 因此不发生缺货现象。

这个模型可用图 8.2 表示。由图 8.2 得知, 每补充一批货物 Q , 库存量即刻由零上升至 Q , 然后以 R 的速率均匀消耗掉。库存量沿斜线下降, 一旦库存量降至零时, 立即再次补充 Q , 如此不断重复。

在以上模型的假设条件下, 怎样确定最佳的储存策略? 一般用单位时间内总费用(平均费用)作为策略优劣的评价指标。为了寻找最佳策略, 即确定一个使平均总费用最小的订货量及订货间隔, 首先要导出平均总费用与定货量 Q 或订货间隔时间 t 之间的函数关系。假设:

Q ——每次订货量;

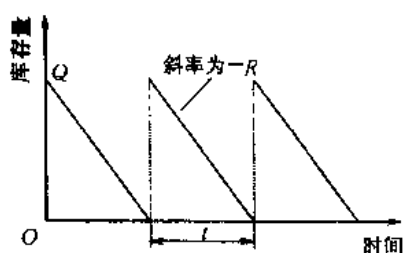


图 8.2

t ——订货间隔时间, $t = \frac{Q}{R}$;

C_3 ——每次订货准备费用;

K ——货物单价;

C_1 ——单位时间内单位货物的储存费。

t 时间内平均库存量为

$$\frac{1}{t} \int_0^t RT dT = \frac{1}{2} R t = \frac{1}{2} Q$$

从而导出作为订货量 Q 的函数的平均总费用

$$C(Q) = (\text{准备费用} + \text{货价} + \text{储存费用}) / \text{时间 } t$$

$$= (C_3 + KQ + C_1 \cdot \frac{1}{2} R t^2) / t$$

$$= \frac{(C_3 + KQ)}{\frac{Q}{R}} + \frac{1}{2} C_1 Q$$

$$= \frac{C_3 R}{Q} + KR + \frac{1}{2} C_1 Q$$

为求 $\min C(Q)$, 求导数

$$\frac{dC(Q)}{dQ} = -\frac{C_3 R}{Q^2} + \frac{C_1}{2} = 0$$

即得最佳定货量

$$Q^* = \sqrt{\frac{2C_3 R}{C_1}}$$

可进一步通过计算 $C(Q)$ 的二阶导数证明 Q^* 是极小点。这就是著名的经济批量 (Economic Ordering Quantity) 公式, 简称 EOQ 公式。由此同时得出最佳订货时间间隔

$$t^* = Q^* / R = \sqrt{\frac{2C_3}{C_1 R}}$$

例 8.1 设某产品年需求量为 144 件, 且需求是均匀的, 每生产一批产品 (不管产量多少) 需准备费用 90 元, 每生产一件产品的直接成本为 100 元, 库存费用为每件每年 20 元。试决定每批生产的最佳数量。

解 需求量 $R = 144$ 件/年, 准备费用 $C_3 = 90$ 元; 生产一件产品直接成本 $K = 100$ 元; 每件每年库存费用 $C_1 = 20$ 元/件年; 每次生产量为 Q 。使用 EOQ 公式, 得

$$Q^* = \sqrt{\frac{2C_3 R}{C_1}} = \sqrt{\frac{2 \times 90 \times 144}{20}} = 36 (\text{件})$$

最低费用 (年平均总费用)

$$\begin{aligned} C(Q) &= \frac{C_3 R}{Q} + KR + \frac{1}{2} C_1 Q \\ &= \frac{90 \times 144}{36} + 100 \times 144 + \frac{1}{2} \times 20 \times 36 \end{aligned}$$

$$= 15\,120 \text{ 元}$$

8.2.2 在制品批量的库存模型(模型 2)

本模型假设条件,除了储存补充是以 P 速度进行外,其余与模型 1 相同。

本模型最典型的例子是描述生产过程中两工序之间需要一定数量的在制品,而上下工序的生产速率并不同步,上工序供应不能发生短缺,才能保证下工序连续进行,这就要求上工序生产速度 P 要大于下工序的需求速度 R 。它的生产过程如图 8.3 描述。开始时上工序以 P 速度生产(即补充),同时下工序以 R 速度消耗在制品,由于 $P > R$,在制品的储存量沿 OA 斜线上升。当储存量达一定数量时,上工序停止生产,下工序仍不断消耗在制品,储存量以 R 速度沿斜线 AB 下降至零,这时上工序即刻开始生产(补充),如此不断重复。

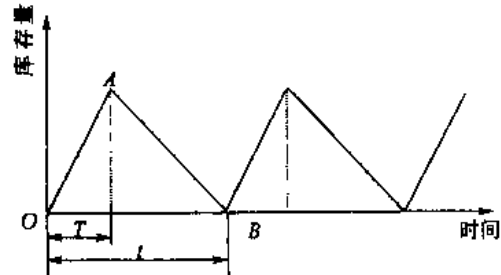


图 8.3

设上工序每次生产时间为 T ,则每次生产批量 $Q = PT$ 。同时为了确定最佳生产批量 Q^* ,首先需导出单位时间内总费用(平均费用)与批量 Q 之间的关系。由图 8.3 分析可知,在 $[0, T]$ 区间内储存量以 $(P - R)$ 速度增长,在 $[T, t]$ 区间内存储量以 R 速度下降,且有

$$(P - R)T = R(t - T)$$

即得

$$T = \frac{Rt}{P}$$

t 时间内平均储存量为 $\frac{1}{t} \left[\frac{1}{2} (P - R) Tt \right]$; t 时间内储存费为 $\frac{1}{2} C_1 (P - R) Tt$; 每生产一次准备成本为 C_3 。

单位时间内总费用(平均费用)

$$\begin{aligned} C(Q) &= \frac{1}{t} \left[C_3 + KQ + \frac{1}{2} C_1 (P - R) Tt \right] \\ &= \frac{R}{Q} \left[C_3 + KQ + \frac{1}{2} C_1 (P - R) \frac{R \cdot \left(\frac{Q}{R}\right)^2}{P} \right] \\ &= \frac{C_3 R}{Q} + KR + \frac{1}{2} C_1 \frac{P - R}{P} Q \end{aligned}$$

进一步计算得

$$\frac{dC(Q)}{dQ} = -\frac{C_3 R}{Q^2} + \frac{1}{2} C_1 \frac{P - R}{P} = 0$$

由此得最佳批量

$$Q^* = \sqrt{\frac{2C_3 R P}{C_1 (P - R)}}$$

从而得最佳生产时间

$$T^* = \frac{Q^*}{P} = \sqrt{\frac{2C_3 R}{C_1 P(P-R)}}$$

$$t^* = \frac{P}{R} T^* = \sqrt{\frac{2C_3 P}{C_1 R(P-R)}}$$

与模型 1 相比较,可知当 $P \rightarrow \infty$ 时, $\frac{P}{P-R} \rightarrow 1$, 模型 2 公式即成模型 1 中的经济批量公式。事实上当生产速度 $P \rightarrow \infty$, 意味着补充速度无限大, 即为瞬时补充, 模型 2 与模型 1 就完全一样了。

例 8.2 某厂每月需甲产品 100 件, 每月生产率为 500 件, 每批装配费为 5 元, 每月每件产品储存费为 0.4 元, 求最佳量及最低费用。

解 已知 $C_3 = 5, C_1 = 0.4, P = 500, R = 100$ 代入公式, 得

$$Q^* = \sqrt{\frac{2C_3 RP}{C_1(P-R)}} = \sqrt{3125} \approx 56 \text{ 件}$$

$$C_0 = \sqrt{\frac{2C_1 C_3 R(P-R)}{P}} \approx 17.89 \text{ 元}$$

8.2.3 允许发生缺货、生产时间很短的库存模型(模型 3)

本模型是在模型 1 的基础上考虑允许发生缺货的情况, 即当库存由于需求而下降到零时, 可以采取等一段时间再订货补充的策略。这样, 在这一段时间内就发生缺货。一方面, 由于需求得不到满足造成经济损失, 要支付缺货费用; 另一方面, 由于采取这种策略, 增长了订货间隔时间, 减少了订货次数, 以至减少准备费用, 同时也减少了平均储存量, 从而减少储存费用, 从总费用角度看, 采取缺货策略可能是有利的。

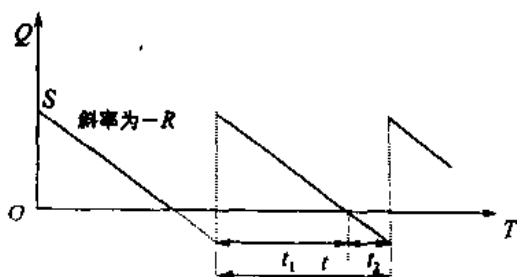


图 8.4

模型 3 可用图 8.4 描述, 其假设条件与模型 1 相同, 并设单位时间内单位缺货费为 C_2 。问题是如何去确定最佳的储存策略, 即不仅要确定最优订货量 Q , 而且要确定最优的拖延订货时间即缺货时间 t_2 , 使得单位时间内总费用(准备费用 + 缺货费用)为最小。分析图 8.4 可看出, 在 t_1 时间间隔内库存量是正的, 随着需求下降为零, 在 t_2 时间间隔

内发生缺货, 最大缺货量为 Rt_2 , 至 t_2 间隔末订货 Q , 补足缺货量 Rt_2 , 库存上升为 $(Q - Rt_2)$, 如此重复。

模型 3 描述的储存系统中, 由于允许缺货, 单位时间内总费用不仅依赖于订货批量(或订货间隔时间), 而且依赖于缺货时间。也就是说, 这平均总费用应是订货量 Q 和缺货时间 t_2 的函数。由图 8.4 可知, 订货间隔时间 $t = \frac{Q}{R}$, $t = t_1 + t_2$; 最大缺货量为 Rt_2 , $Rt_2 = Q - Rt_1$; t 时间内平均储存量为 $\frac{1}{t} [\frac{1}{2} (Q - Rt_2) t_1] = \frac{R}{Q} [\frac{1}{2} (Q - Rt_2)^2]$; t 时间内平均缺货量为 $\frac{1}{t} (\frac{1}{2} Rt_2)$ 。

由此得到单位时间内总费用

$$C(Q, t_2) = \frac{1}{t} [C_3 + \frac{1}{2R} C_1 (Q - Rt_2)^2 + \frac{1}{2} C_2 Rt_2^2]$$

$$= \frac{R}{Q} [C_3 + \frac{1}{2R} C_1 (Q - Rt_2)^2 + \frac{1}{2} C_2 Rt_2^2]$$

由多元函数极值条件, 极小点 (Q^*, t_2^*) 在梯度 $\text{grad } C(Q, t_2) = 0$ 处获得, 即在

$$\frac{\partial C(Q, t_2)}{\partial Q} = 0, \frac{\partial C(Q, t_2)}{\partial t_2} = 0$$

处获得, 则有

$$\frac{\partial C(Q, t_2)}{\partial t_2} = \frac{R}{Q} [-C_1 (Q - Rt_2) + C_2 Rt_2] = 0$$

$$t_2 = \frac{C_1 Q}{(C_1 + C_2)R}$$

$$\frac{\partial C(Q, t_2)}{\partial Q} = -\frac{R}{Q^2} [C_3 + \frac{1}{2R} C_1 (Q - Rt_2)^2 + \frac{1}{2} C_2 Rt_2^2] + \frac{R}{Q} \cdot \frac{C_1}{R} (Q - Rt_2) = 0$$

将 t_2 值代入得最优批量

$$Q^* = \sqrt{\frac{2C_3(C_1 + C_2)R}{C_1 C_2}}$$

从而得

$$t_2^* = \sqrt{\frac{2C_1 C_3}{C_2(C_1 + C_2)R}}$$

$$t^* = \frac{Q^*}{R} = \sqrt{\frac{2C_3(C_1 + C_2)}{C_1 C_2 R}}$$

在上述公式中, 若将条件改为不允许缺货, 即缺货费 $C_2 \rightarrow \infty$, 则 $\frac{C_1 + C_2}{C_2} \rightarrow 1$, 上述公式即趋向模型 1 之 EOQ 公式。

例 8.3 已知需求速度 $R = 100$ 件, $C_1 = 0.4$ 元, $C_2 = 0.15$ 元, $C_3 = 5$ 元, 求最低单位总费用及最佳储存策略下最大库存量。

解 用 $Q^* = \sqrt{\frac{2C_3(C_1 + C_2)R}{C_1 C_2}}$ 及 $t_2^* = \sqrt{\frac{2C_1 C_3}{C_2(C_1 + C_2)R}}$

代入 $C(Q, t_2)$ 中, 得最低单位总费用

$$C(Q^*, t_2^*) = \sqrt{\frac{2C_1 C_2 C_3 R}{C_1 + C_2}} = \sqrt{\frac{2 \times 0.4 \times 0.15 \times 5 \times 100}{0.4 + 0.15}}$$

$$\approx 10.46 (\text{元})$$

最大库存量

$$S = Q^* - t_2^* R$$

$$= \sqrt{\frac{2C_3(C_1 + C_2)R}{C_1 C_2}} - \sqrt{\frac{2C_1 C_3 R}{C_2(C_1 + C_2)}}$$

$$= \sqrt{\frac{2C_2 C_3 R}{C_1(C_1 + C_2)}} = \sqrt{\frac{2 \times 0.15 \times 5 \times 100}{0.4(0.4 + 0.15)}} \approx 26 \text{ 件}$$

8.2.4 允许缺货(需补足缺货)、生产需一定时间的库存模型(模型 4)

假设条件除允许缺货生产需一定时间外,其余条件皆与模型 1 相同。

其存贮变化如图 8.5 所示。

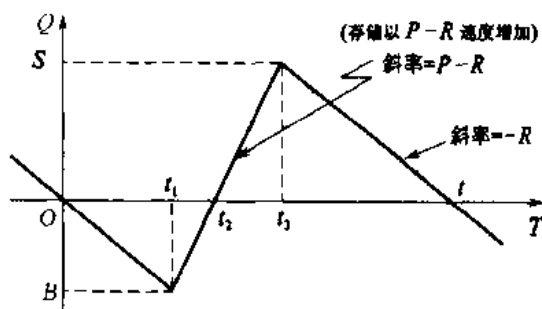


图 8.5

取 $[0, t]$ 为一个周期, 设 t_1 时刻开始生产。

$[0, t_2]$ 时间内存贮为零, B 表示最大缺货量。

$[t_1, t_2]$ 时间内除满足需求外, 还要补足 $[0, t_1]$ 时间内的缺货。

$[t_2, t_3]$ 时间内满足需求后的产品进入存贮, 存贮量以 $(P-R)$ 速度增加。 S 表示存贮量, t_3 时刻存贮量达到最大, t_3 时刻停

止生产。

$[t_3, t]$ 时间存贮量以需求速度 R 减少。

由图 8.5 易知:

最大缺货量 $B = Rt_1$, 或 $B = (P-R)(t_2 - t_1)$ 。即

$$Rt_1 = (P-R)(t_2 - t_1), \text{ 得}$$

$$t_1 = \frac{(P-R)}{P} t_2$$

最大存贮量 $S = (P-R)(t_3 - t_2)$ 或 $S = R(t - t_3)$ 。即 $(P-R)(t_3 - t_2) = R(t - t_3)$, 得

$$t_3 = \frac{R}{P} t + (1 - \frac{R}{P}) t_2, \text{ 或}$$

$$t_3 - t_2 = \frac{R}{P} (t - t_2)$$

在 $[0, t]$ 时间内所需费用:

$$\text{存贮费} \quad \frac{1}{2} C_1 (P-R)(t_3 - t_2)(t - t_2)$$

$$\text{消去 } t_3, \text{ 得} \quad \frac{1}{2} C_1 (P-R) \frac{R}{P} (t - t_2)^2$$

$$\text{缺货费} \quad \frac{1}{2} C_2 R t_1 t_2$$

$$\text{消去 } t_1, \text{ 得} \quad \frac{1}{2} C_2 R \frac{(P-R)}{P} t_2^2$$

$$\text{装配费} \quad C_3$$

在 $[0, t]$ 时间内总平均费用

$$\begin{aligned} C(t, t_2) &= \frac{1}{t} \left[\frac{1}{2} C_1 \frac{(P-R)R}{P} (t - t_2)^2 + \frac{1}{2} C_2 \frac{(P-R)R}{P} t_2^2 + C_3 \right] \\ &= \frac{1}{2} \frac{(P-R)R}{P} \left[C_1 t - 2C_1 t_2 + (C_1 + C_2) \frac{t_2^2}{t} \right] + \frac{C_3}{t} \end{aligned}$$

$$\frac{\partial C(t, t_2)}{\partial t} = \frac{1}{2} \frac{(P-R)R}{P} [C_1 + (C_1 + C_2)t_2^2(-\frac{1}{t^2})] - \frac{C_3}{t^2}$$

$$\frac{\partial C(t, t_2)}{\partial t_2} = \frac{1}{2} \frac{(P-R)R}{P} [-2C_1 + 2(C_1 + C_2)t_2 \cdot \frac{1}{t}]$$

由 $\frac{\partial C(t, t_2)}{\partial t_2} = 0$, 得 $t_2 = \frac{C_1}{C_1 + C_2} t$ 。

由 $\frac{\partial C(t, t_2)}{\partial t} = 0$, 得

$$\frac{1}{2} \frac{(P-R)R}{P} [C_1 + (C_1 + C_2)t_2^2(-\frac{1}{t^2})] - \frac{C_3}{t^2} = 0$$

消去 t_2 得

$$\frac{1}{2} \frac{(P-R)R}{P} [C_1 - (C_1 + C_2) \frac{C_1^2}{(C_1 + C_2)^2}] - C_3 \frac{1}{t^2} = 0$$

$$\frac{1}{2} \frac{(P-R)R}{P} (\frac{C_1^2 + C_1 C_2 - C_1^2}{C_1 + C_2}) - \frac{C_3}{t^2} = 0$$

$$t^2 = \frac{2P(C_1 + C_2)C_3}{C_1 C_2 (P-R)R}$$

则

$$t = \sqrt{\frac{2C_3}{C_1 R}} \cdot \sqrt{\frac{C_1 + C_2}{C_2}} \cdot \sqrt{\frac{P}{P-R}}, \text{记作 } t_0$$

有

$$t_2 = \frac{C_1}{C_1 + C_2} t_0 = \frac{C_1}{C_1 + C_2} \cdot \sqrt{\frac{2C_3}{C_1 R}} \cdot \sqrt{\frac{C_1 + C_2}{C_2}} \cdot \sqrt{\frac{P}{P-R}}$$

依数学分析的知识可以判定 $C(t, t_2)$ 在 $t = t_0, t_2 = \frac{C_1}{C_1 + C_2} t_0$ 时有最小值。

相应地得到

$$t_0 = \sqrt{\frac{2C_3}{C_1 R}} \cdot \sqrt{\frac{C_1 + C_2}{C_2}} \cdot \sqrt{\frac{P}{P-R}}$$

$$Q_0 = R \cdot t_0 = \sqrt{\frac{2C_3 R}{C_1}} \cdot \sqrt{\frac{C_1 + C_2}{C_2}} \cdot \sqrt{\frac{P}{P-R}}$$

$$S_0 (\text{最大存贮量}) = R(t_0 - t_3) = R(t_0 - \frac{R}{P}t_0 - \frac{P-R}{P}t_2)$$

$$= R[t_0 - \frac{R}{P}t_0 - \frac{(P-R)}{P} \cdot \frac{C_1}{(C_1 + C_2)} t_0]$$

$$= R \cdot \frac{(P-R)}{P} \cdot \frac{C_2}{(C_1 + C_2)} \cdot t_0$$

$$= \sqrt{\frac{2C_3 R}{C_1}} \cdot \sqrt{\frac{C_2}{C_1 + C_2}} \cdot \sqrt{\frac{P-R}{P}}$$

$$B_0 (\text{最大缺货量}) = R \cdot t_1 = \frac{R(P-R)}{P} \cdot t_2$$

$$= \sqrt{\frac{2C_1 C_3 R}{(C_1 + C_2) C_2}} \cdot \sqrt{\frac{P-R}{P}}$$

$$\min C(t_0, t_2) = C_0 = \sqrt{2C_1 C_3 R} \cdot \sqrt{\frac{C_2}{C_1 + C_2}} \cdot \sqrt{\frac{P-R}{P}}$$

8.3 随机型库存模型

8.2 节讨论的库存模型中的参数,即需求量及提前期,都是确定的。这一节讨论的几个模型的参数是随机的。

8.3.1 单时期无准备成本的库存模型

所谓单时期模型是指订货库存物资只满足一个特定时期需求的模型。这一节中只讨论这种单时期模型。先假设需求量是随机的,需求是瞬时的,补充也是瞬时的,而且无准备成本。在这些条件下,根据期望总成本最小的原则,去确定最佳储存策略,即确定最佳定货量 Q^* 。

1. 需求是连续的随机变量

一些参数的符号如下:

I ——初始库存量;

Q ——订货量;

y ——订货补充后库存量, $y = Q + I$;

C_1 ——这个时期内单位货物储存费;

C_2 ——这个时期内单位货物缺货费;

K ——单位货价;

ξ ——需求量;

$\varphi(\xi)$ ——需求量 ξ 之概率密度函数。

为了讨论方便,用最佳订货补充后库存量 y^* 代替 Q^* , $y^* = I + Q^*$ 。确定了 y^* 也就确定了 Q^* 。因此首先要导出期望成本与 y 之间的函数关系。由分析知,若补充后库存量为 y 时,这时期储存量

$$H(y) = \begin{cases} y - \xi, & \xi < y \\ 0, & \xi \geq y \end{cases}$$

而缺货量

$$G(y) = \begin{cases} 0, & \xi < y \\ \xi - y, & \xi \geq y \end{cases}$$

则这个时期的期望成本费用

$$\begin{aligned} E[C(y)] &= \text{定货成本} + E[\text{储存费}] + E[\text{缺货费}] \\ &= K(y - I) + C_1 \int_0^\infty H(y) \varphi(\xi) d\xi + C_2 \int_0^\infty G(y) \varphi(\xi) d\xi \\ &= K(y - I) + C_1 \int_0^y (y - \xi) \varphi(\xi) d\xi + C_2 \int_y^\infty (\xi - y) \varphi(\xi) d\xi \end{aligned}$$

求解

$$\frac{dE[C(y)]}{dy} = K + C_1 \int_0^y \varphi(\xi) d\xi - C_2 \int_y^\infty \varphi(\xi) d\xi = 0$$

并利用

$$\int_y^\infty \varphi(\xi) d\xi = 1 - \int_0^y \varphi(\xi) d\xi$$

即得方程

$$\int_0^y \varphi(\xi) d\xi = \frac{C_2 - K}{C_1 + C_2}$$

由此解得满足方程的 y^* 即为所求。(由于 $E[C(y)]$ 之二阶导数 $\frac{d^2 E[C(y)]}{dy^2} = (C_1 + C_2) \varphi(y) > 0$, $E[C(y)]$ 为凸函数, 有惟一最小点 y)

按上述推导, 确定最优 y^* , 即确定 y^* 值使 $\xi < y^*$ 的概率满足

$$P(\xi < y) = \frac{C_2 - K}{C_1 + C_2} = q$$

并称 q 值为临界值。

例 8.4 已知单位储存费 $C_1 = 0.5$ 元, 单位货价 $K = 0.5$ 元, 单位缺货费 $C_2 = 4.5$ 元, 初始库存量 $I = 0$, 需求量密度函数是

$$\varphi(\xi) = \begin{cases} \frac{1}{10}, & 0 \leq \xi \leq 10 \\ 0, & \xi > 10 \end{cases}$$

试求单时期模型最佳订货量。

解 先计算

$$q = \frac{C_2 - K}{C_1 + C_2} = \frac{4.5 - 0.5}{4.5 + 0.5} = 0.8$$

$$P(\xi \leq y) = \int_0^y \frac{1}{10} d\xi = \frac{1}{10} y = 0.8$$

$$y^* = 8$$

$$Q^* = y^* - I = 8 - 0 = 8$$

2. 需求是随机离散的

如果需求量 ξ 是离散的, 则上述期望成本费用

$$E[C(y)] = K(y - I) + C_1 \sum_{\xi=0}^y (y - \xi) \varphi(\xi) + C_2 \sum_{\xi=y+1}^{\infty} (\xi - y) \varphi(\xi)$$

在离散情况下, 可由下列不等式组确定最小点 y^* :

$$E[C(y-1)] \geq E[C(y)] \text{ 和 } E[C(y+1)] \geq E[C(y)]$$

$$\begin{aligned} E[C(y-1)] &= K(y-1-I) + C_1 \sum_{\xi=0}^{y-1} (y-1-\xi) \varphi(\xi) + C_2 \sum_{\xi=y}^{\infty} (\xi - y + 1) \varphi(\xi) \\ &= K(y-I) + C_1 \sum_{\xi=0}^{y-1} (y-\xi) \varphi(\xi) + C_2 \sum_{\xi=y}^{\infty} (\xi - y) \varphi(\xi) - C_1 \sum_{\xi=0}^{y-1} \varphi(\xi) \\ &\quad + C_2 \sum_{\xi=y}^{\infty} \varphi(\xi) - K \end{aligned}$$

$$= E[C(y)] + C_2 - K - (C_1 + C_2) \sum_{\xi=0}^{y-1} \varphi(\xi)$$

进一步得

$$E[C(y-1)] - E[C(y)] = C_2 - K - (C_1 + C_2)P(\xi \leq y-1) \geq 0$$

即

$$P(\xi \leq y-1) \leq \frac{C_2 - K}{C_1 + C_2}$$

同样可由

$$E[C(y+1)] > E[C(y)]$$

得

$$P(\xi \leq y) \geq \frac{C_2 - K}{C_1 + C_2}$$

由此得最小点 y^* 必满足

$$\sum_{\xi=0}^{y^*-1} \varphi(\xi) \leq \frac{C_2 - K}{C_1 + C_2} \leq \sum_{\xi=0}^{y^*} \varphi(\xi)$$

例 8.5(报童问题) 报童每天能售出的报数是个随机变量,每售出一份报能得利润 s 元。当天报未能售出,可退回报社,但每份报损失 h 元;若当天报不够卖,必须从其他报贩处购进报以保证供应,每份报批发价为 K 元。根据以往经验每日能售出报数 ξ 的概率为 $P(\xi)$,问报童每日应订购多少份报纸为最佳?

解 这显然是个单时期无准备成本储存模型。设每日订购报数为 Q 份,则 $I=0, y=Q$ 。可从两个角度来讨论:一是以支付期望总费用最少准则确定 Q^* ;另一是以利润损失期望最小为准则确定 Q^* 。先以期望总费用最少准则出发,确定最佳订货量 Q^* ,有 $Q=y$,则进一步得

$$\begin{aligned} \text{期望总费用 } E[C(y)] = & \text{订报费用} - \text{退回当日售不完报的回收费用} \\ & + \text{从其他报贩购入当日缺货量的费用} \end{aligned}$$

$$E[C(y)] = Ky - (K-h) \sum_{\xi=0}^y (y-\xi)P(\xi) + (K+s) \sum_{\xi=y+1}^{\infty} (\xi-y)P(\xi)$$

相应于 $C_1 = h - K, C_2 = K + s$, 求 $E[C(y)]$ 之最小值,即求解不等式

$$\sum_{\xi=0}^{y-1} P(\xi) \leq \frac{C_2 - K}{C_1 + C_2} = \frac{s}{h+s} \leq \sum_{\xi=0}^y P(\xi)$$

若已知 $s=7, h=4, \xi$ 概率分布如表 8.1 所示。

表 8.1

ξ (千)	0	1	2	3	4	5
$P(\xi)$	0.05	0.10	0.25	0.35	0.15	0.10

先计算临界值 $q = \frac{s}{h+s} = \frac{7}{4+7} = 0.637$, 进一步得

$$\sum_{\xi=0}^2 P(\xi) = 0.4 < 0.637 < \sum_{\xi=0}^3 P(\xi) = 0.75$$

所以, $y^* = 3$, 即每日订货 3 000 张为最佳。

同样,也可以从利润损失最少的角度出发来确定最佳订货量 Q^* 。

期望利润损失 $E[C(y)] = \text{利润损失} + \text{缺货损失}$

$$= h \cdot \sum_{\xi=0}^y (y - \xi) P(\xi) + s \sum_{\xi=y+1}^{\infty} (\xi - y) P(\xi)$$

由不等式

$$E[C(y-1)] > E[C(y)] \text{ 及 } E[C(y+1)] > E[C(y)]$$

来确定 y^* 值, 类似前面分析得

$$\begin{aligned} E[C(y+1)] - E[C(y)] &= h \sum_{\xi=0}^{y+1} (y+1 - \xi) P(\xi) + s \sum_{\xi=y+2}^{\infty} (\xi - y - 1) P(\xi) \\ &\quad - h \sum_{\xi=0}^y (y - \xi) P(\xi) - s \sum_{\xi=y+1}^{\infty} (\xi - y) P(\xi) \\ &= (h + s) \sum_{\xi=0}^y P(\xi) - s \geq 0 \end{aligned}$$

$$\sum_{\xi=0}^y P(\xi) \geq \frac{s}{h + s}$$

同样可得

$$E[C(y-1)] - E[C(y)] = s - (h + s) \sum_{\xi=0}^{y-1} P(\xi) \geq 0$$

即可由

$$\sum_{\xi=0}^{y-1} P(\xi) \leq \frac{s}{h + s} \leq \sum_{\xi=0}^y P(\xi)$$

从两个不同角度出发确定最佳值 y^* , 结论是一致的。

8.3.2 单时期有准备成本库存模型(s - S 策略)

1. 需求是连续的随机变量

本节将讨论在有准备成本 C_3 条件下的单时期库存模型。在这个条件下储存系统的期望成本(初始库存为 I , 订货量 Q , $y = Q + I$)

$$E[C(y)] = \begin{cases} C_3 + K(y - I) + C_1 \int_0^y (y - \xi) \varphi(\xi) d\xi + C_2 \int_y^{\infty} (\xi - y) \varphi(\xi) d\xi, & \text{当 } y > I \text{ (订购)} \\ C_1 \int_0^I (I - \xi) \varphi(\xi) d\xi + C_2 \int_I^{\infty} (\xi - I) \varphi(\xi) d\xi, & \text{当 } y = I \text{ (不订购)} \end{cases}$$

设

$$L(y) = Ky + C_1 \int_0^y (y - \xi) \varphi(\xi) d\xi + C_2 \int_y^{\infty} (\xi - y) \varphi(\xi) d\xi$$

则有

$$E[C(y)] = \begin{cases} C_3 - KI + L(y), & y > I \text{ (订购)} \\ -KI + L(y), & y = I \text{ (不订购)} \end{cases}$$

在这个模型中, 应该解决两个问题: ①在初始库存 I 条件下, 该不该订货; ②若需订货的话, 应该订多少最合适。对这两个问题的回答, 仍然是根据期望总成本费最小的原则。

采用 s - S 策略, 即根据随机需求量 ξ 确定 s 和 S 值。当期初库存 $I < s$ 时, 采取订货策略, 订货补充库存至 S , $S = I + Q$; 当 $I > s$ 时, 则不订货, 目的是确定 s 和 S 值。

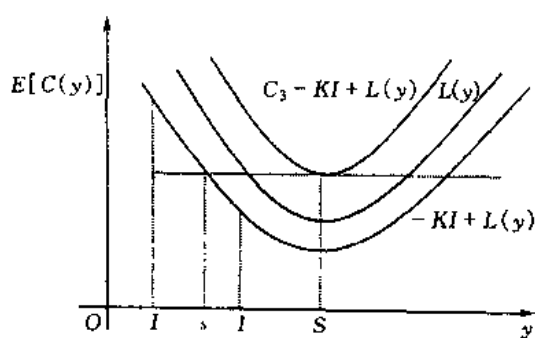


图 8.6

首先确定 S 值。由图 8.6 知, 函数 $C_3 - KI + L(y)$, $-KI + L(y)$ 及函数 $L(y)$ 之极小点 y 是一致的, 取

$$S = y^*$$

即由方程

$$\int_0^y \varphi(\xi) d\xi = \frac{C_2 - K}{C_1 + C_2}$$

确定 y^* , $S = y^*$ 。

其次, 由方程 $L(s) = C_3 + L(S)$ 确定 s

值(同时产生二个 s 值, 舍去 $s > S$ 之值)。然后论证, 这样确定的 s - S 值确实构成最优储存策略。分别从三种情况证明。

①当 $I < s$ 时, 由 $L(y)$ 图形知, $L(I) > L(s)$, 采取不订货策略时期望成本 ($Q = 0, y = I$) $E[C(I)] = -KI + L(I)$; 采取订货策略, 订货补充至 y , 相应期望成本 $E[C(y)] = C_3 - KI + L(y)$ 。考虑其最小期望成本

$$\min_{y>I} E[C(y)] = \min_{y>I} [C_3 - KI + L(y)] = C_3 - KI + L(y^*)$$

其中, y^* 为 $L(y)$ 之最小点, $S = y^*$, 则得

$$\min_{y>I} E[C(y)] = C_3 - KI + L(y^*) = -KI + L(s) < -KI + L(I) = E[C(I)]$$

由此看出, 当 $I < s$ 时, 采取订货策略。订货补充至 S 时, 期望成本比不订货时的期望成本要小, 所以应该采用订货至 S 的策略。

②当 $s < I < S$ 时, 由图 8.6 知 $L(I) < L(s)$ 。若采用订货策略时, 即 $y > I$ 时, 最小期望成本

$$E[C(S)] = \min_{y>I} [C(y)] = C_3 - KI + L(y^*) = C_3 - KI + L(S)$$

若采用不订货策略时, 期望成本

$$E[C(I)] = -KI + L(I) < -KI + L(s) = -KI + C_3 + L(S) = E[C(S)]$$

所以当 $s < I < S$ 时, 应采用不订货策略。

③当 $S < I$ 时, 任何订货使 $y > I$ 时, 显然有

$$E[C(y)] = C_3 - KI + L(y) > -KI + L(I) = E[C(I)]$$

所以不订货为合适。

归纳起来, 对于初始库存为 I 的单时期有准备成本的储存系统, 最优储存策略为:

①当 $I > s$ 时, 不订货;

②当 $I < s$ 时, 订货, 订货量为 $(S - I)$, 其中 S 是函数

$$L(y) = Ky + \int_0^y C_1(y - \xi) \varphi(\xi) d\xi + C_2 \int_y^\infty (\xi - y) \varphi(\xi) d\xi$$

的极小点, 可由方程

$$\int_0^s \varphi(\xi) d\xi = \frac{C_2 - K}{C_1 + C_2}$$

确定。

s 为方程

$$L(s) = C_3 + L(S)$$

的最小值点,显然 s 和 S 与 I 无关,只是与需求量 ξ 之密度函数 $\varphi(\xi)$ 有关。

例 8.6 考虑例 8.4 中,若有准备费用 $C_3 = 9$ 时,试确定最优储存策略 $s-S$ 。

解 由例 8.4 已确定 $y^* = 8$,则取 $S = 8$ 。进一步确定 s 值,考虑函数

$$\begin{aligned} L(y) &= 0.5y + 0.5 \int_0^y \frac{1}{10}(y - \xi) d\xi + 4.5 \int_y^{10} \frac{1}{10}(\xi - y) d\xi \\ &= 0.5y + 0.05(y\xi - \frac{\xi^2}{2}) \Big|_0^y + 0.45(\frac{\xi^2}{2} - y\xi) \Big|_y^{10} \\ &= 0.25y^2 - 4.0y + 22.5 \end{aligned}$$

由方程 $L(s) = C_3 + L(S)$ 得

$$\begin{aligned} 0.25s^2 - 4.0s + 22.5 &= 9 + 0.25S^2 - 4.0S + 22.5 \\ s^2 - 16s + 28 &= 0 \end{aligned}$$

解为

$$s_1 = 2, \quad s_2 = 14 (\text{舍去})$$

最优策略为:当 $I < 2$ 时,应订货,订货量为 $Q^* = 8 - I$;当 $I > 2$ 时,不订货。

2. 需求是随机离散的

当需求量 ξ 是离散取值时,订货补充至 y 时的期望成本

$$E[C(y)] = C_3 + K(y - I) + C_1 \sum_{\xi \leq y} (y - \xi) \varphi(\xi) + C_2 \sum_{y < \xi} (\xi - y) \varphi(\xi)$$

完全类同前面的讨论,最佳值 y^* 可由满足不等式

$$\sum_{\xi=0}^{y^*-1} \varphi(\xi) \leq \frac{C_2 - K}{C_1 + C_2} \leq \sum_{\xi=0}^{y^*} \varphi(\xi)$$

的 y^* 值来确定。取 $S = y^*$,然后同样再确定 s 值。设

$$L(y) = Ky + C_1 \sum_{\xi \leq y} (y - \xi) \varphi(\xi) + C_2 \sum_{\xi > y} (\xi - y) \varphi(\xi)$$

考察不等式

$$L(y) \leq C_3 + L(S)$$

取使不等式成立的最小 y 值为 s 值。

例 8.7 某工厂在某一阶段对原料需求量的概率为: $\varphi(\xi = 80) = 0.1$, $\varphi(\xi = 90) = 0.2$, $\varphi(\xi = 100) = 0.3$, $\varphi(\xi = 110) = 0.3$, $\varphi(\xi = 120) = 0.1$; 订货准备费用 $C_3 = 2825$ 元; 货价 $K = 850$ 元; 单位储存费用 $C_1 = 45$ 元; 单位缺货费用 $C_2 = 1250$ 元。试确定该工厂在这一阶段中的最优储存策略。

解

$$\textcircled{1} \text{ 计算临界值 } q = \frac{C_2 - K}{C_1 + C_2} = \frac{1250 - 850}{1250 + 45} = 0.309$$

② 由计算

$$\varphi(\xi = 80) + \varphi(\xi = 90) = 0.3 < 0.309 < 0.6 = \varphi(\xi = 80) + \varphi(\xi = 90) + \varphi(\xi = 100)$$

确定 $S = 100$ 。

③ 用 $S = 100$ 代入

$$L(s) \leq C_3 + L(S)$$

得

$$L(s) \leq 2825 + (850 \times 100) + 45 \times [(100 - 80) \times 0.1 + (100 - 90) \times 0.2 + (100 - 100) \times 0.3] + 1250 \times [(110 - 100) \times 0.3 + (120 - 100) \times 0.1] = 94255$$

由于 $s < S$, s 只能取 80、90 二值。以 $s = 80$ 代入计算得

$$L(80) = (850 \times 80) + 45 \times (80 - 80) \times 0.1 + 1250 \times [(90 - 80) \times 0.2 + (100 - 80) \times 0.3 + (110 - 80) \times 0.3 + (120 - 80) \times 0.1] = 94250 < 94255$$

所以取 $s = 80$ 。

该厂储存策略应为:当初始库存 $I < 80$ 时,订货 $100 - I$;当 $I > 80$ 时,不订货。

在单时期有准备成本模型中,当需求量 ξ 具有负指数分布时,即

$$\varphi(\xi) = \frac{1}{\lambda} e^{-\frac{\xi}{\lambda}}$$

时,可使用简便的近似方法确定最优储存策略。

确定 S 值时,计算得

$$\int_0^S \varphi(\xi) d\xi = 1 - e^{-\frac{S}{\lambda}} = \frac{C_2 - K}{C_1 + C_2}$$

$$S = \lambda \ln \frac{C_1 + C_2}{C_1 + K}$$

另外有

$$L(y) = Ky + C_1 \int_0^y (y - \xi) \frac{1}{\lambda} e^{-\frac{\xi}{\lambda}} d\xi + C_2 \int_y^\infty (\xi - y) \frac{1}{\lambda} e^{-\frac{\xi}{\lambda}} d\xi$$

$$= (K + C_1)y + \lambda(C_1 + C_2)e^{-\frac{y}{\lambda}} - \lambda C_1$$

为求 s 计算 $L(s) = C_3 + L(S)$, 得

$$(K + C_1)s + \lambda(C_1 + C_2)e^{-\frac{s}{\lambda}} - \lambda C_1 = C_3 + (K + C_1)S + \lambda(C_1 + C_2)e^{-\frac{S}{\lambda}} - \lambda C_1$$

$$\frac{\lambda(C_1 + C_2)}{K + C_1} e^{-\frac{s}{\lambda}} = \frac{C_3}{K + C_1} + (S - s) + \frac{\lambda(C_1 + C_2)}{K + C_1} e^{-\frac{S}{\lambda}}$$

$$e^{\frac{C_3 + C_2}{\ln(K + C_1)} \frac{-s}{\lambda}} = \frac{C_3}{\lambda(K + C_1)} + \frac{S - s}{\lambda} + 1$$

$$e^{\frac{S - s}{\lambda}} = \frac{C_3}{\lambda(K + C_1)} + \frac{S - s}{\lambda} + 1$$

当 $\frac{S - s}{\lambda}$ 很小时,左边按 Taylor 公式展开,舍去二次以后项得

$$1 + \frac{S - s}{\lambda} + \frac{1}{2} \frac{(S - s)^2}{\lambda^2} \approx \frac{C_3}{\lambda(K + C_1)} + \frac{S - s}{\lambda} + 1$$

$$S - s \approx \sqrt{\frac{2\lambda C_3}{K + C_1}}$$

这样得确定 s - S 的近似公式:

$$S = \lambda \ln \frac{C_1 + C_2}{C_1 + K}$$

$$s = S - \sqrt{\frac{2\lambda C_3}{K + C_1}} = \lambda \ln \frac{C_1 + C_2}{C_1 + K} - \sqrt{\frac{2\lambda C_3}{K + C_1}}$$

8.3.3 随机需求与随机补充时间的库存模型

这种模型连续观察库存量,当库存量下降至 R 时,就订货,订货量为 Q 。问题是要确定最优的 R 和 Q 值,使得储存系统在单位时间内期望总费用最小。这一模型假设:发出订单到收到订货之间的提前时间(拖后时间)是随机的;需求量也是随机的。并设:

$r(x/t)$ ——提前时间 t 内需求量为 x 的条件概率密度函数;

$S(t)$ ——提前时间是 $t(t>0)$ 的概率密度函数;

$f(x)$ ——提前时间内需求量为 x 的概率密度函数,且

$$f(x) = \int_0^{\infty} r(x/t) S(t) dt$$

Q ——每个周期的定货量;

D ——每年的期望总需求量;

C_1 ——每年每单位货物储存费;

C_2 ——每次每单位的缺货费用。

每年准备成本 $C_3 \cdot \frac{D}{Q}$ ($\frac{D}{Q}$ 为每年大约订货次数);

每周平均储存量 $H = \frac{(Q + E[R - x]) + E[R - x]}{2}$

$$= \frac{Q}{2} + E[R - x]$$

$$= \frac{Q}{2} + R - \int_0^{\infty} xf(x) dx$$

全年储存费 $C_1 \left[\frac{Q}{2} + R - \int_0^{\infty} xf(x) dx \right];$

每年周期期望缺货量 $\bar{S} = \int_R^{\infty} (x - R) f(x) dx;$

全年缺货费 $C_2 \bar{S} \frac{D}{Q}$

年期望总费用 = 准备成本 + 期望储存费用 + 期望缺货费用,即

$$C(Q, R) = C_3 \frac{D}{Q} + C_1 \left(\frac{Q}{2} + R - E[x] \right) + C_2 \bar{S} \frac{D}{Q}$$

为求最优 Q^* 和 R^* , 计算

$$\begin{cases} \frac{\partial C(Q, R)}{\partial Q} = -\left(\frac{C_3 D}{Q^2}\right) + \frac{C_1}{2} - \frac{C_2 \bar{S} D}{Q^2} = 0 \\ \frac{\partial C(Q, R)}{\partial R} = C_1 - \left(\frac{C_2 D}{Q}\right) \int_R^{\infty} f(x) dx = 0 \end{cases}$$

得

$$\begin{cases} Q^* = \sqrt{\frac{2D(C_3 + C_2 \bar{S})}{C_1}} & (1) \\ \int_{R^*}^{\infty} f(x) dx = \frac{C_1 Q^*}{C_2 D} & (2) \end{cases}$$

求解上述非线性方程组,可采用 Hadley 和 Whitin 提出的方法。只要解存在,此方法便可经有限次迭代收敛。

首先导出解存在的充分条件。令 $R=0$ 时, $\bar{S} = \int_0^{\infty} xf(x)dx = E[x]$, 代入公式(1)得

$$Q^* = \hat{Q} = \sqrt{\frac{2D(C_3 + C_2 E[x])}{C_1}}$$

由公式(2)得

$$Q^* = \tilde{Q} = \frac{C_2 D}{C_1}$$

可以证明,若 $\tilde{Q} \geq \hat{Q}$ 时,满足式(1)和式(2)的 Q^* 的 R^* 存在且惟一。

Hadley-Whitin 方法步骤如下:

以 $\bar{S}_0 = 0, Q_0 = \sqrt{\frac{2DC_3}{C_1}}$ 作 Q^* 之初始值,代入式(2)得 R_0 。由 R_0 计算 \bar{S} 值,得 \bar{S}_1 ,代入式(1)得 Q_1 ,重复,以 Q_1 代入式(2)得 R_1 。以此类推,直至相邻 R_i 和 R_{i+1} 近似相等,则最后得到的 Q_{i+1} 和 R_{i+1} 就近作为 Q^* 和 R^* 。

例 8.8 某储存系统 $C_3 = 100, D = 1\ 000, C_1 = 2, C_2 = 10$, 提前时间内需求量服从一个从 0 到 100 范围内的均匀分布,试确定最优储存策略。

解 首先判断是否有解,取

$$\hat{Q} = \sqrt{\frac{2D(C_3 + C_2 E[x])}{C_1}} = \sqrt{\frac{2 \times 1\ 000 \times (100 + 10 \times 50)}{2}} = 774.5$$

$$\tilde{Q} = \frac{C_2 D}{C_1} = \frac{10 \times 1\ 000}{2} = 5\ 000$$

由于 $\tilde{Q} > \hat{Q}$, 则存在惟一的 Q^* 和 R^* , 期望缺货量

$$\bar{S} = \int_R^{\infty} (x - R)f(x)dx = \int_R^{100} (x - R)\frac{1}{100}dx = \frac{R^2}{200} - R + 50 \quad (3)$$

根据已知条件及式(1)、式(2)得

$$Q^* = \sqrt{\frac{2D(C_3 + C_2 \bar{S})}{C_1}} = \sqrt{\frac{2 \times 1\ 000(100 + 10\bar{S})}{2}} \quad (4)$$

$$\int_R^{100} \frac{1}{100}dx = \frac{2Q^*}{10 \times 1\ 000} \quad (\text{或 } R^* = 100 - \frac{Q^*}{50}) \quad (5)$$

将确定的 Q_i 值代入式(5)得 R_i 值,以 R_i 值代入式(3)和式(4)得 Q_{i+1} , 即有

$$\text{第一次迭代} \begin{cases} Q_1 = \sqrt{\frac{2DC_3}{C_1}} = \sqrt{\frac{2 \times 1\ 000 \times 100}{2}} = 316 \\ R_1 = 100 - \frac{316}{50} = 93.68 \end{cases}$$

$$\text{第二次迭代} \begin{cases} \bar{S} = \frac{R_1^2}{200} - R_1 + 50 = 0.199\ 71 \\ Q_2 = \sqrt{100\ 000 + 10\ 000 \times 0.199\ 71} = 319.37 \\ R_2 = 100 - \frac{319.37}{50} = 93.612 \end{cases}$$

$$\begin{cases} \bar{S} = \frac{R_2^2}{200} - R_2 + 50 = 0.204\ 03 \\ Q_3 = \sqrt{100\ 000 + 10\ 000 \times 0.204\ 03} = 319.43 \\ R_3 = 100 - \frac{319.43}{50} = 93.611 \end{cases}$$

由于 $R_2 \approx R_3$, 则最优解取 $R^* = 93.61$, $Q^* = 319.4$, 即当库存量降至 93.61 时发出订单, 订货量为 319.4。

习 题

8.1 设某工厂每年需用某种原料 1 800 t, 不需每日供应, 但不得缺货。设每吨每月的保管费为 60 元, 每次订购费为 200 元, 试求最佳定购量。

8.2 某公司采用无安全存量的存贮策略。每年使用某种零件 100 000 件, 每件每年的保管费用为 3 元, 每次订购费为 60 元, 试求: (1) 经济定购批量; (2) 如每次订购费为 0.60 元, 每次订购多少?

8.3 设某工厂生产某种零件, 每年需要量为 18 000 个, 该厂每月可生产 3 000 个, 每次生产的装配费为 500 元, 每个零件的存贮费为 0.15 元, 求每次生产的最佳批量。

8.4 某产品每月用量为 4 件, 装配费为 50 元, 存贮费每月每件为 8 元, 求产品每次最佳生产量及最小费用。若生产速度为每月 10 件, 求每次生产量及最小费用。

8.5 每月需要某种机械零件 2 000 件, 每件成本 150 元, 每年的存贮费用为成本的 16%, 每次订购费 100 元, 求 EOQ 及最小费用。

8.6 在题 8.5 中如允许缺货, 求库存量 s 及最大缺货量, 设缺货费 $C_2 = 200$ 元。

8.7 某制造厂每周购进某种机械零件 50 件, 订购费为 4 元, 每周保管费为 0.36 元。 (1) 求 EOQ。 (2) 该厂为少占用流动资金, 希望存贮量达到最低限度, 决定宁可使总费用超过最低费用的 4% 作为存贮策略, 问这时订购批量为多少?

8.8 某公司经理一贯采用不允许缺货的经济批量公式确定订货批量, 因为他认为缺货后补上总不是好事。但由于激烈竞争迫使他不得不考虑采用允许缺货的策略。已知对该公司所销产品的需求 $R = 800$ 件/年, 每次订货费用 $C_1 = 150$ 元, 存贮费 $C_1 = 3$ 元/(件·年), 发生短缺时的损失 $C_2 = 20$ 元/(件·年)。试分析: (1) 计算采用允许缺货的策略较之原先不允许缺货策略带来的费用上的节约; (2) 如果该公司为保持一定信誉, 自己规定缺货随后补上的数量不超过总量的 15%, 任何一名顾客因供应不及时需等下批货到达补上的时间不得超过 3 周, 问这种情况下, 允许缺货的策略能否被采用?

8.9 某商店准备在新年前订购一批挂历批发出售, 已知每售出一批 (100 本) 可获利 70 元。如果挂历在新年前售不出去, 则每 100 本损失 40 元。根据以往销售经验, 该商店售出挂历数量的概率如表 8.2 所示。如果该商店对挂历只能提出一次订货, 问应订几百本, 使期望的获利数为最大。

表 8.2

销售量(百本)	0	1	2	3	4	5
概 率	0.05	0.10	0.25	0.35	0.15	0.10

8.10 某商店代销一种产品, 每件产品的购进价格为 800 元, 存贮费每件 40 元, 缺货费每件 1 015 元, 订购费一次 60 元, 原有库存 10 件。已知对产品需求的概率见表 8.3。试确定该商店的最佳订货数量。

表 8.3

需求量 x	30	40	50	60
$P(x)$	0.20	0.20	0.40	0.20

8.11 某商店存有某种商品 10 件, 每件的进价为 3 元, 存贮费为 1 元, 缺货费为 16 元。已知对该种商品

的需求量服从 $\mu = 20$ 、 $\sigma = 5$ 的正态分布,求商店对该种商品的最佳订货量。

8.12 某商店准备订购一批圣诞树迎接节日,据历年经验,销量服从正态分布, $\mu = 200$, $\sigma^2 = 300$ 。每棵圣诞树售价为 25 元,进价为 15 元。如果进了货卖不出去,则节后其残值基本为零。试回答:(1)该商店应进多少棵圣诞树,使期望利润值为最大;(2)如果商店按销售量的期望值 200 棵进货,则期望的利润值为多大;(3)如商店按(1)计算数字进货,则未能销售出去的圣诞树的期望值是多少?

第9章 多阶段决策——动态规划

9.1 多阶段决策问题

在生产、计划、管理中往往需要研究处理包含多个阶段决策过程的问题。这类问题能分解为若干阶段或若干子问题,通过对每个子问题作出决策得到解决。动态规划就是研究这种多阶段决策问题的优化方法。它能为分析问题的全过程提供总的框架,在这个框架内又可用各种优化技术解决每一阶段上的具体问题。

动态规划的理论基础是由 R. Bellman 提出的最优性原理。这个原理归结为一个递推关系表达式,用它描述多阶段决策过程的状态转移。一般采用逆序方法求这类问题的解,即从最终状态出发,逐步推算到初始状态,从而得到一个最优决策序列。与其他优化技术相比,动态规划能得到一个多阶段的整体最优解。

因为决策过程的时间参数可连续或离散,故决策过程可分为连续决策过程和离散决策过程。根据决策过程演变是确定性的还是随机性的,决策过程又可分为确定性决策过程和随机性决策过程。综合起来便有离散确定性、离散随机性、连续确定性、连续随机性共四种决策过程模型。

本书主要阐述离散决策过程,介绍动态规划的基本概念、理论及方法,并以各种实例说明其应用。

多阶段决策问题很多,下面首先通过两个具体例子说明什么是多阶段决策问题。

例 9.1 最短路问题:现欲从图

9.1 中的 A 点沿道路铺设管道至 E 点,应如何铺设管道,使得管子总长度最短?

这个问题从图形看就是求 A 至 E 的最短路径问题,可以看成是一个多阶段决策问题。把它分成 4 个阶段考虑,每个阶段需从它所在位置选择这一阶段铺设的道路,连接各个阶段选择的通路即是 A 至 E 的一条路径。这样构成的路径有许多条,问题是如何在每个阶段作出适当选择,而使构成的路径为最短。

例 9.2 机器负荷分配问题:某种机器可以在高低两种不同负荷下生产。在高负荷下生产时,产品的年产量 y_1 和投入生产机器数 x 的关系为

$$y_1 = g(x)$$

这时机器完好率为 a ,即如果年初完好机器数为 x ,到年终完好机器数为 $ax, 0 < a < 1$;在低负

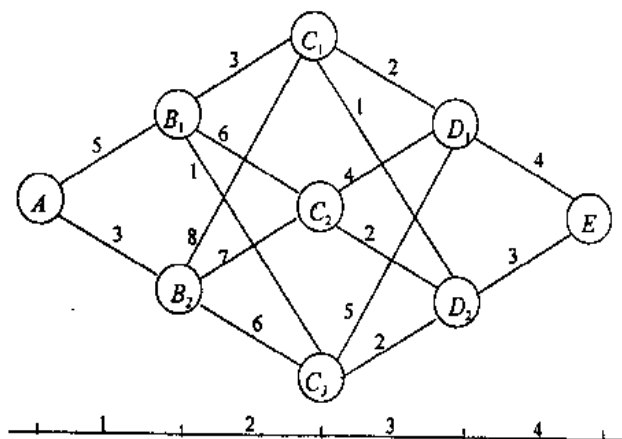


图 9.1

荷生产时,产品的年产量 y_2 和投入生产的机器数量 x 的关系为

$$y_2 = f(x)$$

相应机器完好率为 $b, 0 < b < 1$ 。

假定开始生产时完好机器数为 S_1 。试制定一个五年计划,在每年初决定重新分配在两种负荷下生产的机器数,使五年内总产量达最高?

这是一个典型的多阶段决策问题,每年为一个阶段,每阶段需决定投入多少台机器在高负荷下生产,使得5年总产量最高(图9.2)。

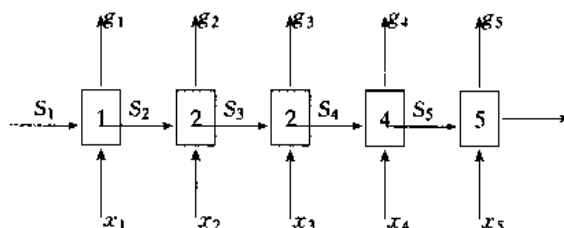


图 9.2

由上述实例可知,多阶段决策问题是这样一种特殊的决策过程,即可将该过程分为若干个互相联系的阶段,在它的每一阶段都需作出决策,从而使整个过程达到最优。当然,各个阶段决策的选取不是任意确定的,它依赖于当前面临的状态,又影响以后的发展。当各个阶段决策确定后,就构成了一个决策序列,因而也就决定了整个过程的一条活动路线。这种把一个问题看做是一个前后关联具有链状结构的多阶段过程就称为多阶段决策过程,也称序贯决策过程,相应的问题就称为多阶段决策问题。

9.2 动态规划的基本概念

用动态规划处理多阶段决策问题时,首先得建立一些基本概念,通过这些概念定量描述这个多阶段决策问题。现在结合例9.1解释这些概念。

9.2.1 阶段与阶段变量

原问题分成若干相互联系的子问题,每个子问题称为一个阶段,用变量 k 表示, $k = 1, 2, \dots, n$, 称其为阶段变量。如由图9.1所示的最短路问题就分成4个阶段, $k = 1, 2, 3, 4$ 。

9.2.2 状态与状态变量

过程各阶段所处的“位置”称为状态,用变量 S_k 表示 k 阶段初所处的状态,称其为状态变量。某阶段所有可能状态的全体称为状态集合。例9.1中, S_3 表示第三阶段初起点之位置,状态集合 $\{S_3\} = \{C_1, C_2, C_3\}$ 。例9.2中, S_2 表示第二年初完好机器的数目。

9.2.3 决策与决策变量

某阶段初状态决定后,从这状态向下一阶段哪个状态演变的选择称为决策,用变量 $x_k = x_k(S_k)$ 表示由 k 阶段 S_k 状态出发所作的决策,称之为决策变量。决策变量 $x_k(S_k)$ 可选择的范围称为允许决策集合,用 $D(x_k)$ 表示。

9.2.4 状态转移

前一阶段的状态和决策决定了下一阶段的状态,它们之间的关系称为状态转移,用方程

$$S_{k+1} = T[S_k, x_k(S_k)]$$

表示。如图 9.2 中, S_k 表示 k 年初完好机器数(状态); $x_k(S_k)$ 表示决定投入高负荷生产的机器数(决策),则下一年初完好的机器数

$$S_{k+1} = ax_k(S_k) + b(S_k - x_k(S_k))$$

即是由 S_k 状态经决策 $x_k(S_k)$ 至 S_{k+1} 状态的转移方程。

9.2.5 策略

由阶段 $k=1$ 至阶段 $k=n$ 的全过程中,由每个阶段所选择的决策构成一决策序列,称之为一个策略,记为

$$P_1(S_1) = \{x_1(S_1), x_2(S_2), \dots, x_n(S_n)\}$$

在例 9.1 中每一策略即表示由 $A \rightarrow E$ 的一条通路,如 $P_1(S_1) = \{A, B_1, C_2, D_2, E\}$ 表示是一个策略。

9.2.6 后部子策略

从 k 阶段某状态 S_k 出发到终点的过程称为后部过程。它相应的决策序列称为后部子策略,记为

$$P_{kn}(S_k) = \{x_k(S_k), \dots, x_n(S_n)\}$$

如在例 9.1 中,从 $k=2, S_2=B_1$ 出发的一个后部子策略

$$P_{2n}(S_2) = \{B_1, C_2, D_2, E\}$$

即表示由 B_1 至 E 的一条通路。

9.2.7 阶段指标

k 阶段中,评价由状态 S_k 出发作出决策 $x_k(S_k)$ 所产生效果的数量指标记为

$$V_k(S_k, x_k)$$

在例 9.1 中,如由 $k=2, S_2=B_1$ 出发,选择决策 $x_2(B_1)=C_1$,则阶段指标 $V_2(B_1, C_1)=2$,即用 $B_1 C_1$ 之路长作为决策效果评价的数量指标。

9.2.8 指标函数与最优指标函数

评价由 k 阶段 S_k 状态出发到终点的后部子策略 $P_{kn}(S_k)$ 产生效果的数量指标记为

$$V_{kn}[S_k, P_{kn}(S_k)] = V_{kn}(S_k, x_k, S_{k+1}, x_{k+1}, \dots, S_n, x_n)$$

它是用来衡量所实现过程优劣的一种数量指标,这个数量指标由状态 S_k 和后部子策略 $P_{kn}(S_k)$ 决定,所以称为 S_k 和 $P_{kn}(S_k)$ 的指标函数。而由 k 阶段 S_k 状态出发,所有可能的后部子策略所产生的指标函数值中最优者称为最优指标函数值,记为 $f_k(S_k)$,即

$$f_k(S_k) = \underset{\{x_k, \dots, x_n\}}{\text{opt}} V_{kn}[S_k, P_{kn}(S_k)]$$

显然它是由状态 S_k 决定,称为状态 S_k 的最优指标函数。特别当 $k=1$ 时, $f_1(S_1)$ 表示全过程的最优指标值,相应的策略 $P_{1n}(S_1)$ 即为最优策略。在例 9.1 中,当 $k=2, S_2=C_1$ 时,每一个后部策略表示由 $C_1 \rightarrow E$ 的一条路径,相应指标函数值就是这条路的路长。而最优指标函数值 $f_3(C_1)$ 为 $C_1 \rightarrow E$ 所有通路中最短的路长。 $f_1(A)$ 表示由 $A \rightarrow E$ 的最短路长,相应最优策略 $P_{1n}^*(A)$ 为 $A \rightarrow E$ 的最短路径。

9.3 最优性原理与递推方程

本节通过例 9.1 说明动态规划怎样根据最优性原理,通过递推形式解决多阶段决策问题。动态规划的理论依据是 Bellman 提出的最优性原理。

最优性原理 一个过程的最优策略具有这样的性质,即无论过去状态和决策如何,对前面决策形成的状态而言,余下的决策必构成最优策略。这就是最优性原理。

这个原理实际上是常识性的。如在例 9.1 中,若

$$A \rightarrow B_1 \rightarrow C_3 \rightarrow D_2 \rightarrow E$$

是 $A \rightarrow E$ 的最优策略(最短路),则 $C_3 \rightarrow D_2 \rightarrow E$ 必是 $C_3 \rightarrow E$ 的最短路径。

现在看例 9.1。例 9.1 的问题是寻找 $A \rightarrow E$ 的最短路径(最优策略)及最短路长(最优指标函数值 $f_1(A)$)。如果能求出

$$\begin{cases} B_1 \rightarrow E \\ B_2 \rightarrow E \end{cases} \text{的最短路及最短路长} \begin{cases} f_2(B_1) \\ f_2(B_2) \end{cases}$$

则根据 Bellman 最优性原理,相应能求出 $A \rightarrow E$ 的最优指标函数值 $f_1(A)$ 及最短路,即有

$$f_1(A) = \min \{ V_1(A, B_1) + f_2(B_1), V_1(A, B_2) + f_2(B_2) \}$$

其中 $V_1(A, B_1)$ 为 AB_1 的路长, $V_1(A, B_2)$ 为 AB_2 的路长。

这样,求 $A \rightarrow E$ 最短路问题 1 转为求 $B_1 \rightarrow E$ 和 $B_2 \rightarrow E$ 最短路问题 2。

同样,欲求 $B_1 \rightarrow E$ 最短路问题,可转化为求

$$\begin{cases} C_1 \rightarrow E \\ C_2 \rightarrow E \\ C_3 \rightarrow E \end{cases} \text{最短路及最短路长} \begin{cases} f_3(C_1) \\ f_3(C_2) \\ f_3(C_3) \end{cases}$$

$$f_2(B_1) = \min \begin{cases} V_2(B_1, C_1) + f_3(C_1) \\ V_2(B_1, C_2) + f_3(C_2) \\ V_2(B_1, C_3) + f_3(C_3) \end{cases}$$

同样,求 $B_2 \rightarrow E$ 最短路问题有

$$f_2(B_2) = \min \begin{cases} V_2(B_2, C_1) + f_3(C_1) \\ V_2(B_2, C_2) + f_3(C_2) \\ V_2(B_2, C_3) + f_3(C_3) \end{cases}$$

依此类推,求 $C_1 \rightarrow E$ 最短路问题 3 转为求

$$\begin{cases} D_1 \rightarrow E \\ D_2 \rightarrow E \end{cases} \text{的最短路及最短路长} \begin{cases} f_4(D_1) \\ f_4(D_2) \end{cases}$$

的问题 4, 这样由逆向递推求解

$$\begin{array}{lcl}
 \left\{ \begin{array}{l} D_1 \\ D_2 \end{array} \right\} \rightarrow E \text{ 最短路} & \Rightarrow & \left\{ \begin{array}{l} C_1 \\ C_2 \\ C_3 \end{array} \right\} \rightarrow E \text{ 最短路} \\
 \text{(问题 4)} & & \text{(问题 3)} \\
 \rightarrow \left\{ \begin{array}{l} B_1 \\ B_2 \end{array} \right\} \rightarrow E \text{ 最短路} & \Rightarrow & A \rightarrow E \text{ 最短路} \\
 \text{(问题 2)} & & \text{(问题 1)}
 \end{array}$$

而问题 4 求 $D_i \rightarrow$ 最短路是容易解得的。整个求解过程由下列步骤表示:

$$k=4, f_4(D_1) = V_4(D_1, E) = 4$$

$$f_4(D_2) = V_4(D_2, E) = 3$$

$$k=3, f_3(C_1) = \min \left\{ \begin{array}{l} V_3(C_1, D_1) + f_4(D_1) \\ V_3(C_1, D_2) + f_4(D_2) \end{array} \right\} = \min \left\{ \begin{array}{l} 2+4 \\ 1+3 \end{array} \right\} = 4$$

$$f_3(C_2) = \min \left\{ \begin{array}{l} V_3(C_2, D_1) + f_4(D_1) \\ V_3(C_2, D_2) + f_4(D_2) \end{array} \right\} = \min \left\{ \begin{array}{l} 4+4 \\ 2+3 \end{array} \right\} = 5$$

$$f_3(C_3) = \min \left\{ \begin{array}{l} V_3(C_3, D_1) + f_4(D_1) \\ V_3(C_3, D_2) + f_4(D_2) \end{array} \right\} = \min \left\{ \begin{array}{l} 5+4 \\ 2+3 \end{array} \right\} = 5$$

$$k=2, f_2(B_1) = \min \left\{ \begin{array}{l} V_2(B_1, C_1) + f_3(C_1) \\ V_2(B_1, C_2) + f_3(C_2) \\ V_2(B_1, C_3) + f_3(C_3) \end{array} \right\} = \min \left\{ \begin{array}{l} 3+4 \\ 6+5 \\ 1+5 \end{array} \right\} = 6$$

$$f_2(B_2) = \min \left\{ \begin{array}{l} V_2(B_2, C_1) + f_3(C_1) \\ V_2(B_2, C_2) + f_3(C_2) \\ V_2(B_2, C_3) + f_3(C_3) \end{array} \right\} = \min \left\{ \begin{array}{l} 8+4 \\ 7+5 \\ 6+5 \end{array} \right\} = 11$$

$$k=1, f_1(A) = \min \left\{ \begin{array}{l} V_1(A, B_1) + f_2(B_1) \\ V_1(A, B_2) + f_2(B_2) \end{array} \right\} = \min \left\{ \begin{array}{l} 5+6 \\ 3+11 \end{array} \right\} = 11$$

则 $A \rightarrow E$ 最短路长为 $f_1(A) = 11$, 最优策略即最短路可由反向追踪得

$$A \rightarrow B_1 \rightarrow C_3 \rightarrow D_2 \rightarrow E$$

从上述求解过程可看出, 动态规划是通过求解递推过程

$$\begin{cases} f_k(S_k) = \text{opt}[V_k(S_k, x_k) + f_{k+1}(S_{k+1})] \\ x_k \in D(x_k) \quad k = n, (n-1), \dots, 1 \\ f_{n+1}(S_{n+1}) = 0 \end{cases} \quad (9.1)$$

递推求得最优策略 $P_{1n}^*(S_1)$ 及最优指标函数。式(9.1)称为动态规划的递推方程或基本方程。其中, $D(x_k)$ 为允许决策集合, $S_{k+1} = T(S_k, x_k)$, 称 $f_{n+1}(S_{n+1}) = 0$ 为边界条件, “opt”是英文最优化的缩写, 可根据题意取 min 或 max。

确切地讲, 式(9.1)只是动态规划逆序解法的基本方程。相应地, 可以给出顺序解法的基本方程。所谓顺序解法是指寻优途径是从初始阶段开始到最后阶段结束即恰与逆序解法的寻优方向相反的一种动态规划问题的求解方法。

以下给出动态规则顺序解法的基本方程如下:

$$\begin{cases} f_k(S_{k+1}) = \text{opt}[V_k(S_{k+1}, x_k) + f_{k-1}(S_k)] \\ x_k \in D'(x_{k+1}) \quad k=1, 2, \dots, n \\ f_0(S_1) = 0 \end{cases} \quad (9.2)$$

上式中,阶段变量 k 及状态变量 S_k 的定义不变,决策变量 x_k 的定义发生变化,即这时的状态转移不是由 S_k, x_k 去确定 S_{k+1} ,而是反过来由 S_{k+1}, x_k 去确定 S_k ,则状态转移方程一般形式为 $S_k = T'_k(S_{k+1}, x_k)$,因而第 k 阶段允许决策集合也相应改变,记为 $D'(x_{k+1})$ 。其求解过程是根据边界条件,从 $k=1$ 开始,从前向后顺推,逐步可求得各阶段的最优决策和相应的最优值,最后求出 $f_n(S_{n+1})$ 时,就得到整个问题的最优解。

本书主要阐述利用逆序解法求解动态规划问题的方法。

最后,可将动态规划解法的基本步骤归纳如下:

①规定过程的分段,构造出状态,定义状态变量,使状态序列能描述整个过程,状态具有无后效性;(所谓“无后效性”是指:若某阶段状态给定,则在这阶段以后过程的发展只与当前的状态有关,而与过去历史无关。)

②定义决策变量,确定允许决策集合;

③确定状态转移规律,写出转移方程;

④定出指标函数,写出递推方程,求解。

对于离散的状态变量及决策变量问题,常采用列表形式递推计算 $f_k(S_k)$ 及 $x_k^*(S_k)$,通常有表 9.1 和表 9.2 两种表格形式。

表 9.1 k 阶段[计算 $f_k(S_k)$ 及 $x_k^*(S_k)$]表

S_k	$V_k(S_k, x_k) + f_{k+1}(S_{k+1})$				$f_k(S_k)$	x_k^*
	x_{k1}	x_{k2}	x_{knt}		

表 9.2 k 阶段[计算 $f_k(S_k)$ 及 $x_k^*(S_k)$]表

S_k	x_k	$V_k(S_k, x_k)$	$S_{k+1} = T(S_k, x_k)$	$f_{k+1}(S_{k+1})$	$V_k(S_k, x_k) + f_{k+1}(S_{k+1})$	$f_k(S_k)$	x_k^*

9.4 动态规划应用例解

例 9.3(资源分配问题) 某工厂有 100 台机器,拟分四期使用,在每一周期有两种生产任务。根据经验要把机器 x_1 投入第一种生产任务,则在一个生产周期中将有 $\frac{1}{3}x_1$ 台机器报废,余下机器全部投入第二种生产任务,则在一个生产周期中将有 $\frac{1}{10}$ 台机器报废。如果在一生产

周期中干第一种生产任务每一台机器可收益 10,干第二种生产任务每一台机器可收益 7。问怎样分配机器使总收益最大?

解 这可视为一个 4 阶段决策问题,可用动态规划方法求解。设阶段变量 $k=1,2,3,4$ 。

状态变量 S_k 表示 k 年初完好机器台数。决策变量 x_k 表示 k 年初决定投入第一种生产任务的机器数。状态转移方程为

$$S_{k+1} = \frac{2}{3}x_k + \frac{9}{10}(S_k - x_k) \quad (0 \leq x_k \leq S_k)$$

阶段指标

$$V(S_k, x_k) = 10x_k + 7(S_k - x_k)$$

$$k=4, f_4(S_4) = \max_{0 \leq x_4 \leq S_4} \{10x_4 + 7(S_4 - x_4) + f_5(S_5)\} = \max_{0 \leq x_4 \leq S_4} \{3x_4 + 7S_4\} \\ = 10S_4 \text{ (取 } x_4^* = S_4 \text{)}$$

$$k=3, f_3(S_3) = \max_{0 \leq x_3 \leq S_3} \{10x_3 + 7(S_3 - x_3) + f_4(S_4)\} \\ = \max_{0 \leq x_3 \leq S_3} \left\{ 10x_3 + 7(S_3 - x_3) + 10 \left[\frac{2}{3}x_3 + \frac{9}{10}(S_3 - x_3) \right] \right\} \\ = \max_{0 \leq x_3 \leq S_3} \left\{ \frac{2}{3}x_3 + 16S_3 \right\} = \frac{50}{3}S_3 \text{ (取 } x_3^* = S_3 \text{)}$$

$$k=2, f_2(S_2) = \max_{0 \leq x_2 \leq S_2} \{10x_2 + 7(S_2 - x_2) + f_3(S_3)\} \\ = \max_{0 \leq x_2 \leq S_2} \left\{ 10x_2 + 7(S_2 - x_2) + \frac{50}{3} \left[\frac{2}{3}x_2 + \frac{9}{10}(S_2 - x_2) \right] \right\} \\ = \max_{0 \leq x_2 \leq S_2} \left\{ 22S_2 - \frac{8}{9}x_2 \right\} = 22S_2 \text{ (取 } x_2^* = 0 \text{)}$$

$$k=1, f_1(S_1) = \max_{0 \leq x_1 \leq S_1} \{10x_1 + 7(S_1 - x_1) + f_2(S_2)\} \\ = \max_{0 \leq x_1 \leq S_1} \left\{ 10x_1 + 7(S_1 - x_1) + 22 \left[\frac{2}{3}x_1 + \frac{9}{10}(S_1 - x_1) \right] \right\} \\ = \max_{0 \leq x_1 \leq S_1} \{26.8S_1 - 2.13x_1\} = 26.8S_1 \text{ (取 } x_1^* = 0 \text{)}$$

最优策略即各期机器最优分配如下:

$$S_1 = 100, x_1^* = 0; \quad S_2 = 90, x_2^* = 0$$

$$S_3 = 81, x_3^* = 81; \quad S_4 = 54, x_4^* = 54$$

最大总收益 $f_1(100) = 2680$ 。

例 9.4(生产计划问题) 某公司与一顾客订立合同,在 4 个月内出售一定数量的某种产品。由于各种原因,每月至多生产 100 单位,产量限于 10 的倍数。产品可以贮存,贮存费用每单位 2 元。生产成本及每月销售额如表 9.3 给出。要求确定一个生产过程,使能满足合同要求,在生产能力以内使生产成本最小。

表 9.3

月 份	单位生产成本 C_i (元)	合同销售额 d_i
1	70	60
2	72	70
3	80	120
4	76	60

解 阶段变量 $k=1,2,3,4$ 表示月份。状态变量 S_k 表示 k 月初已有产品数。决策变量 x_k 表示决定 k 月的生产数量,满足约束

$$x_k + S_k \geq d_k, x_k \leq 100, x_k = 0, 10, 20, \dots$$

状态转移 $S_{k+1} = S_k + x_k - d_k$, 阶段指标 $V_k(S_k, x_k) = c_k x_k + 2S_k$ 。

$k=4$ (表 9.4)时,由于 1~3 月份最大生产量为 300 单位,合同销售总额为 250 单位,所以 4 月份最大贮存量为 50 单位,即 S_4 可能取值为 0、10、20、30、40、50。求解

$$f_4(S_4) = \min_{\substack{0 \leq x_4 \leq 100 \\ x_4 + S_4 \geq d_4}} (C_4 x_4 + 2S_4)$$

得 $x_4^* = d_4 - S_4 = 60 - S_4$, 则有

$$f_4(S_4) = 4560 - 74S_4$$

表 9.4

S_4	$f_4(S_4)$	x_4^*
0	4560	60
10	3820	50
20	3080	40
30	2340	30
40	1600	20
50	860	10

$k=3$ (表 9.5)时,第一、第二月最大生产量为 200 单位,销售合同额为 $60 + 70 = 130$,所以 3 月初最大贮存量为 70 单位。由 $x_3 + S_3 \geq d_3 = 120$ 和 $x_3 \leq 100$ 得

$$S_3 \geq 20$$

所以 S_3 可能取值为 20、30、40、50、60、70 单位。求解

$$f_3(S_3) = \min_{\substack{0 \leq x_3 \leq 100 \\ x_3 + S_3 \geq d_3 = 120}} \{C_3 x_3 + 2S_3 + f_4(S_4)\}$$

表 9.5

$S_3 \backslash x_3$	$V_3(S_3, x_3) + f_4(S_4)$						$f_3(S_3)$	x_3^*
	50	60	70	80	90	100		
20						12 600	12 600	100
30					11 820	11 880	11 820	90
40				11 040	11 100	11 160	11 040	80
50			10 260	10 320	10 380	10 440	10 260	70
60		9 480	9 540	9 600	9 660	9 720	9 480	60
70	8 700	8 760	8 820	8 880	8 940	9 000	8 700	50

$k=2$ (表 9.6)时,1 月份最大生产量为 100 单位,合同销售量为 60,则 2 月初最大贮存量为 $100 - 60 = 40$,即 S_2 可能取值 0、10、20、30、40。求解

$$f_2(S_2) = \min_{\substack{0 \leq x_2 \leq 100 \\ x_2 \geq d_2 - S_2}} \{C_2 x_2 + 2S_2 + f_3(S_3)\}$$

又由 $S_3 = S_2 + x_2 - d_2 \geq 20$, 得

$$x_2 \geq d_2 - S_2 + 20 = 90 - S_2$$

表 9.6

$x_2 \backslash S_2$	$V_2(S_2, x_2) + f_3(S_3)$						$f_2(S_2)$	x_2^*
	50	60	70	80	90	100		
0					19 080	19 020	19 020	100
10				18 380	18 320	18 260	18 260	100
20			17 680	17 620	17 560	17 500	17 500	100
30		16 980	16 920	16 860	16 800	16 740	16 740	100
40	16 280	16 220	16 160	16 100	16 040	15 980	15 980	100

$k=1$ (表 9.7)时, $S_1=0$, 且有

$$f_1(S_1) = \min_{\substack{0 \leq x_1 \leq 100 \\ x_1 + S_1 \geq d_1 = 60 (x_1 \geq 60)}} \{C_1 x_1 + 2S_1 + f_2(S_2)\}$$

表 9.7

$x_1 \backslash S_1$	$V_1(S_1, x_1) + f_2(S_2)$					$f_1(S_1)$	x_1^*
	60	70	80	90	100		
0	23 220	23 160	23 100	23 040	22 980	22 980	100

最小总成本 $f_1(0) = 22\,980$, 最优生产安排如表 9.8 所示。

表 9.8

月份(i)	S_i	x_i	d_i	$c_i x_i$	$2S_i$	$V_i(S_i, x_i)$
1	0	100	60	7 000	0	7 000
2	40	100	70	7 200	80	7 280
3	70	50	120	4 000	140	4 140
4	0	60	60	4 560	0	4 500

例 9.5(投资分配) 现有 8 百万资金欲向三个企业($i=1,2,3$)投资, $V_i(x)$ ($i=1,2,3$)是向第 i 个企业投资 x 时产生的利润函数, 其数据如表 9.9 所示。试求使总利润最大的投资计划。

表 9.9

x	0	1	2	3	4	5	6	7	8
$V_1(x)$	0	5	15	40	80	90	95	98	100
$V_2(x)$	0	5	15	40	60	70	73	74	75
$V_3(x)$	0	4	26	40	45	50	51	52	53

解 可把这个问题作为一个多阶段决策问题来考虑, 如图 9.3 所示。阶段变量 $k=1,2,3$ 。

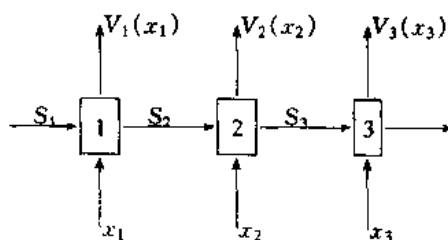


图 9.3

状态变量 $S_1 = 8$ 表示可向 1、2、3 企业投资的资金总数, S_2 表示可向 2、3 企业投资的资金; S_3 表示可向 3 企业投资的资金。

决策变量 x_k 表示对 k 企业的投资数($0 \leq x_k \leq S_k$)。状态转移方程为

$$S_{k+1} = S_k - x_k$$

递推方程为

$$\begin{cases} f_k(S_k) = \max_{0 \leq x_k \leq S_k} \{V_k(x_k) + f_{k+1}(S_{k+1})\} & k=3,2,1 \\ f_4(S_4) = 0 \end{cases}$$

逆推的过程($k=3,2,1$)如表 9.10~9.12 所示。

表 9.10

S_3	x_3	$V_3(x_3)$	S_4	$f_4(S_4)$	$V_3(x_3) + f_4(S_4)$	$f_3(S_3)$	x_3^*
0	0	0	0	0	0	0	0
1	1	4	0	0	4	4	1
2	2	26	0	0	26	26	2
3	3	40	0	0	40	40	3
4	4	45	0	0	45	45	4
5	5	50	0	0	50	50	5
6	6	51	0	0	51	51	6
7	7	52	0	0	52	52	7
8	8	53	0	0	53	53	8

表 9.11

S_2	x_2	$V_2(x_2)$	S_3	$f_3(S_3)$	$V_2(x_2) + f_3(S_3)$	$f_2(S_2)$	x_2^*
	0	0	0	0	0	0	0
1	0	0	1	4	4	5	1
	1	5	0	0	5		
2	0	0	2	26	26	26	0
	1	5	1	4	9		
	2	15	0	0	15		
3	0	0	3	40	40	40	0,3
	1	5	2	26	31		
	2	15	1	4	19		
	3	40	0	0	40		
4	0	0	4	45	45	60	4
	1	5	3	40	45		
	2	15	2	26	41		
	3	40	1	4	44		
	4	60	0	0	60		
5	0	0	5	50	50	70	5
	1	5	4	45	50		
	2	15	3	40	55		
	3	40	2	26	66		
	4	60	1	4	64		
	5	70	0	0	70		

续表

S_2	x_2	$V_2(x_2)$	S_3	$f_3(S_3)$	$V_2(x_2) + f_3(S_3)$	$f_2(S_2)$	x_2^*
6	0	0	6	51	51	86	4
	1	5	5	50	55		
	2	15	4	45	60		
	3	40	3	40	80		
	4	60	2	26	86		
	5	70	1	4	74		
	6	73	0	0	73		
7	0	0	7	52	52	100	4
	1	5	6	51	56		
	2	15	5	50	65		
	3	40	4	45	85		
	4	60	3	40	100		
	5	70	2	26	96		
	6	73	1	4	77		
8	7	74	0	0	74	110	5
	0	0	8	53	53		
	1	5	7	52	57		
	2	15	6	51	66		
	3	40	5	50	90		
	4	60	4	45	105		
	5	70	3	40	110		
	6	73	2	26	99		
	7	74	1	4	78		
	8	75	0	0	75		

表 9.12

S_1	x_1	$V_1(x_1)$	S_2	$f_2(S_2)$	$V_1(x_1) + f_2(S_2)$	$f_1(S_1)$	x_1^*
8	0	0	8	110	110	140	4
	1	5	7	100	105		
	2	15	6	86	101		
	3	40	5	70	110		
	4	80	4	60	140		
	5	90	3	40	130		
	6	95	2	26	121		
	7	98	1	5	103		
	8	100	0	0	100		

最大总利润 $f_1(8) = 140$

最优投资策略:

$$S_1 = 8, x_1^* = 4 \quad S_2 = 4, x_2^* = 4 \quad S_3 = 0, x_3^* = 0$$

即最优投资计划为向第一个企业投资 400 万元,向第二个企业投资 400 万元,向第三个企业投资为零。

例 9.6(更新问题) 某运输公司 1985 年初购进一批已使用二年的卡车。假设每年的各

项费用如表 9.13 所示。考虑 1986~1989 年间,为使总利润最大,每年年初对这批车应作怎样的决策(更新或维修)。

表 9.13

使用年数		0	1	2	3	4	5	6
1985	收 入	14		10				
	维修费	1		3				
	更新费	20		25				
1986	收 入	16	16		8			
	维修费	1	1		3			
	更新费	20	22		26			
1987	收 入	18	14	16		8		
	维修费	1	1	2		4		
	更新费	20	22	24		27		
1988	收 入	18	16	14	14		6	
	维修费	1	1	2	2		4	
	更新费	21	22	24	25		28	
1989	收 入	20	16	16	12	12		4
	维修费	1	1	2	2	3		5
	更新费	21	22	24	25	26		29

解 用动态规划法求解。假设用 $R_k(S_k)$ 表示 k 年初,车龄为 S_k 的卡车继续使用一年的收入;用 $U_k(S_k)$ 表示 k 年初,车龄为 S_k 的卡车继续使用一年所需的维修费;用 $C_k(S_k)$ 表示 k 年初,车龄为 S_k 的卡车当年更新所需费用。显然, $R_k(0)$ 表示 k 年期间新车使用一年的收入, $U_k(0)$ 表示 k 年期间新车使用一年的维修费。阶段变量 $k=1,2,3,4,5$, 表示 1985、1986、1987、1988、1989 年。

状态变量 S_k 表示 k 年初卡车的车龄。决策变量 $x_k=0$ 表示更新, $x_k=1$ 表示继续使用。状态转移方程为

$$S_{k+1} = x_k \cdot S_k + 1$$

状态集合

$$\{S_k\} = \{1, 2, \dots, k-1, k+1\}$$

阶段指标(年收入)

$$V_k(S_k, x_k) = R_k(x_k, S_k) - U_k(x_k, S_k) - (1 - x_k)C_k(S_k)$$

递推方程为

$$\begin{cases} f_k(S_k) = \max \{ R_k(x_k, S_k) - U_k(x_k, S_k) - (1 - x_k)C_k(S_k) + f_{k+1}(x_k S_k + 1) \} \\ x_k = 0, 1 \\ f_6(S_6) = 0, k = 5, 4, 3, 2, 1 \end{cases}$$

递推过程如表 9.14~9.18 所示。

表 9.14

S_5	x_5	$S_6 = x_5 S_5 + 1$	$V_5(S_5, x_5) + f_6(S_6)$	$f_5(S_5)$	x_5^*
1	0	1	$20 - 1 - 22 + 0$	15	1
	1	2	$16 - 1 + 0$		
2	0	1	$20 - 1 - 24 + 0$	14	1
	1	3	$16 - 2 + 0$		
3	0	1	$20 - 1 - 25 + 0$	10	1
	1	4	$12 - 2$		
4	0	1	$20 - 1 - 26 + 0$	9	1
	1	5	$12 - 3$		
6	0	1	$20 - 1 - 29 + 0$	-1	1
	1	7	$4 - 5$		

表 9.15

S_4	x_4	$S_5 = x_4 S_4 + 1$	$V_4(S_4, x_4) + f_5(S_5)$	$f_4(S_4)$	x_4^*
1	0	1	$18 - 1 - 22 + 15$	29	1
	1	2	$16 - 1 + 14$		
2	0	1	$18 - 1 - 24 + 15$	22	1
	1	3	$14 - 2 + 10$		
3	0	1	$18 - 1 - 25 + 15$	21	1
	1	4	$14 - 2 + 9$		
5	0	1	$18 - 1 - 28 + 15$	4	0
	1	6	$6 - 4 - 1$		

表 9.16

S_3	x_3	$S_4 = x_3 S_3 + 1$	$V_3(S_3, x_3) + f_4(S_4)$	$f_3(S_3)$	x_3^*
1	0	1	$18 - 1 - 22 + 29$	35	1
	1	2	$14 - 1 + 22$		
2	0	1	$18 - 1 - 24 + 29$	35	1
	1	3	$16 - 2 + 21$		
4	0	1	$18 - 1 - 27 + 29$	19	0
	1	5	$8 - 4 + 4$		

表 9.17

S_2	x_2	$S_3 = x_2 S_2 + 1$	$V_2(S_2, x_2) + f_3(S_3)$	$f_2(S_2)$	x_2^*
1	0	1	$16 - 1 - 22 + 35$	50	1
	1	2	$16 - 1 + 35$		
3	0	1	$16 - 1 - 26 + 35$	24	0.1
	1	4	$8 - 3 + 19$		

表 9.18

S_1	x_1	$S_2 = x_1 S_1 + 1$	$V_1(S_1, x_1) + f_2(S_2)$	$f_1(S_1)$	x_1^*
2	0	1	$14 - 1 - 25 + 50$	38	0
	1	3	$10 - 3 + 24$		

五年内最优安排为

1985 年初 $S_1 = 2, x_1 = 0$ 更新;

1986 年初 $S_2 = 1, x_2 = 1$ 继续使用;

1987 年初 $S_3 = 2, x_3 = 1$ 继续使用;

1988 年初 $S_4 = 3, x_4 = 1$ 继续使用;

1989 年初 $S_5 = 4, x_5 = 1$ 继续使用。

五年内最大利润为 $f_1(2) = 38$ 。

例 9.7(可靠性问题) 考虑设计一种有三个主要元件的电气设备。这三个元件是串联的, 所以一个元件故障将使整个设备发生故障。设备的可靠性(不发生故障的概率)可以通过在每个元件上并联(备用的)元件改进, 每个元件可以由一个、二个或三个元件并联。设计这种设备的总资金为 10(单位: 千元)。对元件 $i(i=1, 2, 3)$ 配备 m_i 个元件($m_i=1, 2, 3$)后的可靠性为 $R_i(m_i)$, 成本为 $C_i(m_i)$, 具体数字由表 9.19 给出。试确定各元件并联的元件数目, 使得在不超过总设计费用条件下, 使设备可靠性最大。

表 9.19

m_i	$i=1$		$i=2$		$i=3$	
	R	C	R	C	R	C
1	0.6	1	0.7	3	0.5	2
2	0.8	2	0.8	5	0.7	4
3	0.9	3	0.9	6	0.9	5

解 设备总可靠性 $R = R_1(m_1)R_2(m_2)R_3(m_3)$, 其中 $R_i(m_i)$ 是第 i 个元件有 m_i 个并联元件时的可靠性。

阶段变量 $k=1, 2, 3$, 表示元件号。状态变量 S_k 表示设计 k 号元件前剩下的资金数。决策变量 x_k 表示选择 k 号元件的并联数。状态转移方程为

$$S_{k+1} = S_k - C_k(x_k)$$

递推方程为

$$\begin{cases} f_k(S_k) = \max \{ R_k(x_k) f_{k+1}(S_{k+1}) \} & x_k = 1, 2, 3 \\ C_k(x_k) \leq S_k - \sum_{i=k+1}^3 C_i(1) \\ f_4(S_4) = 1 & k = 3, 2, 1 \end{cases}$$

考虑 $k=3$ 时, 满足

$$\begin{cases} C_3(1) \leq S_3 \leq S_1 - C_1(1) - C_2(1) & (2 \leq S_3 \leq 6) \\ C_3(x_3) \leq S_3, x_3 = 1, 2, 3 \end{cases}$$

迭代过程如表 9.20 所示。

表 9.20

S_3	x_3	$R_3(x_3)$	$S_4 = S_3 - C_3(x_3)$	$f_4(S_4)$	$R_3(x_3) \cdot f_4(S_4)$	$f_3(S_3)$	x_3^*
2	1	0.5	0	1	0.5	0.5	1
3	1	0.5	1	1	0.5	0.5	1
4	1	0.5	2	1	0.5	0.7	2
	2	0.7	0	1	0.7		
5	1	0.5	3	1	0.5	0.9	3
	2	0.7	1	1	0.7		
	3	0.9	0	1	0.9		
6	1	0.5	4	1	0.5	0.9	3
	2	0.7	2	1	0.7		
	3	0.9	1	1	0.9		

当 $k=2$ 时, 满足

$$\begin{cases} C_3(1) + C_2(1) \leq S_2 \leq S_1 - C_1(1) & (5 \leq S_2 \leq 9) \\ C_2(x_2) \leq S_2 - C_3(1), x_2 = 1, 2, 3 \end{cases}$$

迭代过程如表 9.21 所示。

表 9.21

S_2	x_2	$R_2(x_2)$	$S_3 = S_2 - C_2(x_2)$	$f_3(S_3)$	$R_2(x_2) \cdot f_3(S_3)$	$f_2(S_2)$	x_2^*
5	1	0.7	2	0.5	0.35	0.35	1
6	1	0.7	3	0.5	0.35	0.35	1
7	1	0.7	4	0.7	0.49	0.49	1
	2	0.8	2	0.5	0.40		
8	1	0.7	5	0.9	0.63	0.63	1
	2	0.8	3	0.5	0.40		
	3	0.9	2	0.5	0.45		
9	1	0.7	6	0.9	0.63	0.63	1
	2	0.8	4	0.7	0.56		
	3	0.9	3	0.5	0.45		

当 $k=1$ 时, 满足

$$\begin{cases} C_1(x_1) \leq S_1 - C_2(1) - C_3(1) = 6 \\ x_i = 1, 2, 3 \end{cases}$$

迭代过程如表 9.22 所示。

表 9.22

S_1	x_1	$R_1(x_1)$	$S_2 = S_1 - C_1(x_1)$	$f_2(S_2)$	$R_1(x_1) \cdot f_2(S_2)$	$f_1(S_1)$	x_1^*
10	1	0.6	9	0.63	0.378	0.504	2
	2	0.8	8	0.63	0.504		
	3	0.9	7	0.49	0.441		

最优安排为

$$S_1 = 10, x_1 = 2 \quad S_2 = 8, x_2 = 1 \quad S_3 = 5, x_3 = 3$$

即 1 号元件并联 2 个, 2 号元件并联 1 个, 3 号元件并联 3 个。最优可靠性 $f_1(10) = 0.504$ 。

例 9.8(背包问题) 一艘货船欲去装运 N 种货物, 第 i 种货物的单位重量为 W_i , 价值为

$V_i (i=1,2,\dots,N)$ 。船的最大装载重量是 W 。现在要确定在不超过船的最大载重条件下,如何装载这 N 种货物,使得船上货物的价值最大。特别考虑下面只有三种货物的特殊情况,并假定 $W=5, W_i$ 和 V_i 数据由表 9.23 给出。

表 9.23

货 号 i	单位重量 W_i	单位价值 V_i
1	2	65
2	3	80
3	1	30

解 使用动态规划求解。

阶段变量 $k (k=1,2,\dots,N)$ 表示货号。状态变量 S_k 表示装 k 号货前船的剩余吨位。决策变量 x_k 表示装运 k 号货物的件数。决策变量允许集合

$$D(x_k) = \{0, 1, 2, \dots, [S_k/W_k]\}$$

状态转移方程

$$S_{k+1} = S_k - x_k W_k$$

阶段指标 $V_k(S_k, x_k) = V_k x_k$ 表示装 k 号货的价值。

递推方程为

$$\begin{cases} f_k(S_k) = \max_{x_k} \{V_k x_k + f_{k+1}(S_k - x_k W_k)\} \\ f_{N+1}(S_{N+1}) = 0 \quad k = N, N-1, \dots, 1 \end{cases}$$

考虑仅有三种货物情况:

①当 $k=3$ 时, $f_3(S_3) = \max\{30x_3\}$, 其中, $x_3 \in \{0, 1, 2, \dots, [S_3]\}$, 如表 9.24 所示;

表 9.24

S_3	x_3	$V_3(S_3, x_3) = 30x_3$	$S_4 = S_3 - x_3$	$f_4(S_4)$	$V_3 + f_4(S_4)$	$f_3(S_3)$	x_3^*
0	0	0	0	0	0	0	0
1	0	0	1	0	0	30	1
	1	30	0	0	30		
2	0	0	2	0	0	60	2
	1	30	1	0	30		
	2	60	0	0	60		
3	0	0	3	0	0	90	3
	1	30	2	0	30		
	2	60	1	0	60		
	3	90	0	0	90		
4	0	0	4	0	0	120	4
	1	30	3	0	30		
	2	60	2	0	60		
	3	90	1	0	90		
	4	120	0	0	120		

续表

S_3	x_3	$V_3(S_3, x_3) = 30x_3$	$S_4 = S_3 - x_3$	$f_4(S_4)$	$V_3 + f_4(S_4)$	$f_3(S_3)$	x_3^*
5	0	0	5	0	0	150	5
	1	30	4	0	30		
	2	60	3	0	60		
	3	90	2	0	90		
	4	120	1	0	120		
	5	150	0	0	150		

②当 $k=2$ 时, $f_2(S_2) = \max\{80x_2 + f_3(S_2 - 3x_2)\}$, 其中, $x_2 \in \{0, 1, 2, \dots, [S_2/3]\}$, 如表 9.25 所示;

表 9.25

S_2	x_2	$V_2(S_2, x_2) = 80x_2$	$S_3 = S_2 - 3x_2$	$f_3(S_3)$	$V_2 + f_3(S_3)$	$f_2(S_2)$	x_2^*
0	0	0	0	0	0	0	0
1	0		1	30	30	30	0
2	0	0	2	60	60	60	0
3	0	0	3	90	90	90	0
	1	80	0	0	80		
4	0	0	4	120	120	120	0
	1	80	1	30	110		
5	0	0	5	150	150	150	0
	1	80	2	60	140		

③当 $k=1$ 时, $f_1(S_1) = \max\{65x_1 + f_2(S_1 - 2x_1)\}$, 其中, $x_1 \in \{0, 1, 2, \dots, [S_1/2]\}$, 如表 9.26 所示。

表 9.26

S_1	x_1	$V_1(S_1, x_1) = 65x_1$	$S_2 = S_1 - 2x_1$	$f_2(S_2)$	$V_1 + f_2(S_2)$	$f_1(S_1)$	x_1^*
5	0	0	5	150	150	160	2
	1	65	3	90	155		
	2	130	1	30	160		

最优装货方案:

$$S_1 = 5, x_1 = 2 \quad S_2 = 1, x_2 = 0 \quad S_3 = 1, x_3 = 1$$

即第一种货装 2 件, 第二种货不装, 第三种货装 1 件, 最大价值为 $f_1(5) = 160$ 。

例 9.9(静态规划问题的动态规划解法) 线性规划与非线性规划所研究的问题通常是与时间无关的, 常称它们为静态规划。对于某些静态规划问题, 也可以人为地引入时间因素, 把它看做是按阶段进行的一个动态规划问题, 这就使得动态规划成为求解一些线性、非线性规划的有效方法。下面分别给出用逆序解法和顺序解法求解静态规划的例子。

①用逆序解法求解下列静态规划问题:

$$\max z = 2x_1^2 + 9x_2 + 4x_3$$

$$\text{s. t. } \begin{cases} x_1 + x_2 + x_3 = 10 \\ x_i \geq 0, i = 1, 2, 3 \end{cases}$$

解 阶段 $k = 1, 2, 3$ 。

状态变量为 S_k , 决策变量为 x_k 。状态转移

$$S_1 = 10, S_2 = S_1 - x_1, S_3 = S_2 - x_2, S_4 = S_3 - x_3 = 0$$

阶段指标

$$V_1(x_1) = 2x_1^2, V_2(x_2) = 9x_2, V_3(x_3) = 4x_3$$

递推方程

$$\begin{cases} f_k(S_k) = \max_{x_k} \{V_k(x_k) + f_{k+1}(S_{k+1})\} \\ f_4(S_4) = 0, k = 3, 2, 1 \end{cases}$$

$k=3$, 求解

$$f_3(S_3) = \max_{0 \leq x_3 \leq S_3} \{4x_3\} \text{ 得}$$

$$x_3^* = S_3, f_3(S_3) = 4S_3$$

$k=2$, 求解

$$f_2(S_2) = \max_{0 \leq x_2 \leq S_2} \{9x_2 + 4(S_2 - x_2)\} = \max_{0 \leq x_2 \leq S_2} \{5x_2 + 4S_2\}$$

$$x_2^* = S_2, f_2(S_2) = 9S_2$$

$k=1$, 求解

$$f_1(S_1) = \max_{0 \leq x_1 \leq S_1} \{2x_1^2 + 9(S_1 - x_1)\} = \max_{0 \leq x_1 \leq S_1} \{2x_1^2 - 9x_1 + 9S_1\}$$

得

$$x_1 = S_1 = 10, f_1(S_1) = 2S_1^2 = 200$$

规划问题解为 $S_1 = 10, x_1^* = 10; S_2 = 0, x_2^* = 0; S_3 = 0, x_3^* = 0$ 。即最优解 $x_1^* = 10, x_2^* = 0, x_3^* = 0$, 最优值 $z^* = 200$ 。

②用顺序解法求解下列静态规划问题:

$$\max z = 4x_1^2 - x_2^2 + 2x_3^2 + 12$$

$$\text{s. t. } \begin{cases} 3x_1 + 2x_2 + x_3 \leq 9 \\ x_i \geq 0 \quad (i = 1, 2, 3) \end{cases}$$

用顺序解法与已知初始状态用逆序解法在本质上没有区别。它相当于把实际的起点视为终点, 实际的终点视为起点, 而按逆推解法进行。对于一般的计算步骤, 有兴趣的读者可参看有关文献。以下给出上例的具体求解过程。

解 按问题中变量的个数分为三个阶段。设状态变量为 S_0, S_1, S_2, S_3 , 并记 $S_3 \leq 9$; 取 x_1, x_2, x_3 为各阶段的决策变量; 各阶段指标函数按加法方式结合。令最优值函数 $f_k(S_k)$ 表示第 k 阶段的结束状态为 S_k , 从 1 阶段至 k 阶段的最大值。

设

$$3x_1 = S_1, S_1 + 2x_2 = S_2, S_2 + x_3 = S_3 \leq 9$$

则有

$$x_1 = S_1/3, 0 \leq x_2 \leq S_2/2, 0 \leq x_3 \leq S_3$$

于是用顺推方法, 从前向后依次有

$$f_1(S_1) = \max_{x_1 = S_1/3} (4x_1^2) = \frac{4}{9} S_1^2 \quad \text{及最优解} \quad x_1^* = S_1/3$$

$$\begin{aligned} f_2(S_2) &= \max_{0 \leq x_2 \leq S_2/2} [-x_2^2 + f_1(S_1)] = \max_{0 \leq x_2 \leq S_2/2} \left[-x_2^2 + \frac{4}{9}(S_2 - 2x_2)^2 \right] \\ &= \max_{0 \leq x_2 \leq S_2/2} h_2(S_2, x_2) \end{aligned}$$

$$\text{由 } \frac{dh_2}{dx_2} = \frac{14}{9}x_2 - \frac{16}{9}S_2 = 0, \text{ 解得 } x_2 = \frac{8}{7}S_2。$$

因该点不在允许决策集合内,故无需判别。因而 $h_2(S_2, x_2)$ 的最大值必在两个端点上选取。而

$$h_2(0) = \frac{4}{9}S_2^2, h_2\left(\frac{S_2}{2}\right) = -S_2^2/4$$

所以 $h_2(S_2, x_2)$ 的最大值点在 $x_2 = 0$ 处,故得到 $f_2(S_2) = \frac{4}{9}S_2^2$ 及相应的最优解 $x_2^* = 0$ 。

又有

$$\begin{aligned} f_3(S_3) &= \max_{0 \leq x_3 \leq S_3} [2x_3^2 + 12 + f_2(S_2)] \\ &= \max_{0 \leq x_3 \leq S_3} \left[2x_3^2 + 12 + \frac{4}{9}(S_3 - x_3)^2 \right] \\ &= \max_{0 \leq x_3 \leq S_3} h_3(S_3, x_3) \end{aligned}$$

$$\text{由 } \frac{dh_3}{dx_3} = \frac{44}{9}x_3 - \frac{8}{9}S_3 = 0, \text{ 解得 } x_3 = \frac{2}{11}S_3。 \text{ 又 } \frac{d^2h_3}{dx_3^2} = \frac{44}{9} > 0, \text{ 故该点为极小值点。而}$$

$$h_3(0) = \frac{4}{9}S_3^2 + 12, h_3(S_3) = 2S_3^2 + 12$$

故 $h_3(S_3, x_3)$ 的最大值点在 $x_3 = S_3$ 处。所以得 $f_3(S_3) = 2S_3^2 + 12$ 及相应的最优解 $x_3^* = S_3$ 。

由于 S_3 不知道,故需再对 S_3 求一次极值,即

$$\max_{0 \leq S_3 \leq 9} f_3(S_3) = \max_{0 \leq S_3 \leq 9} [2S_3^2 + 12]$$

显然,当 $S_3 = 9$ 时 $f_3(S_3)$ 才能达到最大值。所以 $f_1(9) = 2 \times 9^2 + 12 = 174$ 为最大值。

再按计算的顺序反推算可求得最优解为 $x_1^* = 0, x_2^* = 0, x_3^* = 9$; 最大值为

$$\max z = f_1(9) = 174$$

现说明以下两点:

①若先作代换,令 $y_1 = 3x_1, y_2 = 2x_2, y_3 = x_3$ 。将原问题变为

$$\begin{aligned} \max z &= \frac{4}{9}y_1^2 - \frac{1}{4}y_2^2 + 2y_3^2 + 12 \\ \text{s.t. } &\begin{cases} y_1 + y_2 + y_3 \leq 9 \\ y_i \geq 0, y = 1, 2, 3 \end{cases} \end{aligned}$$

再解此问题,然后换回 x_i 也可以;

②在计算 h_2 和 h_3 的最大值中,若学习了凸函数的性质,也可利用凸函数的性质来确定最大值。

从本例可见,一般而言,当初始状态给定时,用逆推比较方便。当终止状态给定时,用顺推比较方便。

在实际中,还会遇到某些多阶段决策过程。这种过程不像前面所讨论的确定性那样,状态转移是完全确定的,而是带有随机性,状态转移不能完全确定,它是按某种已知的概率分布取值的。具有这种性质的多阶段决策过程就是本书前面所讲的随机性决策过程。这类问题同样可以用动态规划的方法解决。下面举两个例子。

例 9.10(随机采购问题) 某厂生产上需要在近五周内采购一批原料,而估计在未来五周内价格有波动,浮动价格和概率已测得如表 9.27 所示。试求在哪一周以什么价格购入,可使采购价格的数学期望值最小,并求出期望值。

表 9.27

单 价	概 率
500	0.3
600	0.3
700	0.4

解 这里价格是一个随机变量,是按某种已知的概率分布取值的。用动态规划方法处理,按采购期限 5 周分为 5 个阶段,将每周的价格看做该阶段的状态。现设以下变量:

y_k 为状态变量,表示第 k 周的实际价格;

x_k 为决策变量(当 $x_k = 1$,表示第 k 周决定采购,当 $x_k = 0$,表示第 k 周决定等待);

y_{kE} 表示第 k 周决定等待,而在以后采取最优决策时采购价格的期望值。

$f_k(y_k)$ 表示第 k 周实际价格为 y_k 时,从第 k 周至第 5 周采取最优决策所得的最小期望值,因而递推关系为

$$\begin{aligned} f_k(y_k) &= \min\{y_k, y_{kE}\} \quad (y_k \in S_k) \\ f_5(y_5) &= y_5 \quad (y_5 \in S_5) \end{aligned} \quad (9.3)$$

其中 $S_k = \{500, 600, 700\}$, $k = 1, 2, 3, 4, 5$ 。由 y_{kE} 和 $f_k(y_k)$ 的定义可知

$$y_{kE} = E f_{k+1}(y_{k+1}) = 0.3 f_{k+1}(500) + 0.3 f_{k+1}(600) + 0.4 f_{k+1}(700) \quad (9.4)$$

并且得出最优决策为

$$x_k = \begin{cases} 1 & (\text{采购}), f_k(y_k) = y_k \\ 0 & (\text{等待}), f_k(y_k) = y_{kE} \end{cases} \quad (9.5)$$

具体计算过程如下:

$k = 5$ 时,由 $f_5(y_5) = y_5$ ($y_5 \in S_5$) 有 $f_5(500) = 500$, $f_5(600) = 600$, $f_5(700) = 700$ 。即在第 5 周时,若所需的原料尚未买入,则无论市场价格如何,都必须采购,不能再等。

$k = 4$ 时,由式(9.4)可知

$$y_{4E} = 0.3 \times 500 + 0.3 \times 600 + 0.4 \times 700 = 610$$

于是,由式(9.3)得

$$f_4(y_4) = \min_{y_4 \in S_4} \{y_4, y_{4E}\} = \min_{y_4 \in S_4} \{y_4, 610\} = \begin{cases} 500, y_4 = 500 \\ 600, y_4 = 600 \\ 610, y_4 = 700 \end{cases}$$

由式(9.5)可知,第四周的最优决策为

$$x_4 = \begin{cases} 1 & (\text{采购}), y_4 = 500 \text{ 或 } 600 \\ 0 & (\text{等待}), y_4 = 700 \end{cases}$$

同理求得

$$f_3(y_3) = \min\{y_3, y_{3E}\} = \min\{y_3, 574\} = \begin{cases} 500, & y_3 = 500 \\ 574, & y_3 = 600 \text{ 或 } 700 \end{cases}$$

所以
$$x_3 = \begin{cases} 1, & y_3 = 500; \\ 0, & y_3 = 600 \text{ 或 } 700 \end{cases}$$

$$f_2(y_2) = \min\{y_2, y_{2E}\} = \min\{y_2, 551.8\} = \begin{cases} 500, & y_2 = 500 \\ 551.8, & y_2 = 600 \text{ 或 } 700 \end{cases}$$

所以
$$x_2 = \begin{cases} 1, & y_2 = 500 \\ 0, & y_2 = 600 \text{ 或 } 700 \end{cases}$$

$$f_1(y_1) = \min_{y_1 \in S_1} \{y_1, y_{1E}\} = \min\{y_1, 536.26\} = \begin{cases} 500, & y_1 = 500 \\ 536.26, & y_1 = 600 \text{ 或 } 700 \end{cases}$$

所以

$$x_1 = \begin{cases} 1, & y_1 = 500 \\ 0, & y_1 = 600 \text{ 或 } 700 \end{cases}$$

由上可知,最优采购策略为在第一、二、三周时,若价格为 500 就采购,否则应该等待;在第四周时,价格为 500 或 600 时应采购,否则就等待;在第五周时,无论什么价格都要采购。

依照上述最优策略进行采购时,价格(单价)的数学期望值为

$$\begin{aligned} & 500 \times 0.3[1 + 0.7 + 0.7^2 + 0.7^3 + 0.7^3 \times 0.4] \\ & + 600 \times 0.3[0.7^3 + 0.4 \times 0.7^3] + 700 \times 0.4^2 \times 0.7^3 \\ & = 500 \times 0.80106 + 600 \times 0.14406 + 700 \times 0.05488 \\ & = 525.382 \approx 525 \end{aligned}$$

且 $0.80106 + 0.14406 + 0.05488 = 1$ 。

例 9.11(随机生产计划问题) 某厂和公司订了试制某种新产品的合同。如果三个月生产不出一个合格品,则要罚款 200 元,每次试制的个数不限,试制周期为一个月,制造一个产品的成本为 100 元,每一个试制品合格的概率为 0.4,生产一次的装配费为 200 元。问如何安排试制和每次生产几个,才能使期望费用最小?

解 根据题意,最多能安排三次生产,把三次试制当做三个阶段,每次生产的个数作为决策变量 x_k ,每次试制前是否已有合格品作为状态变量 u_k ,有合格品时记 $u_k = 0$,无合格品时记 $u_k = 1$ 。 $f_k(u_k)$ 为第 k 次试制前的状态为 u_k 时,以后均采用最优策略时的最低期望成本。为简化数字,取百元为单位。

由假设,当 $u_k = 0$ 时,即已有合格品,试制已完成,于是 $f_k(0) = 0$,即不生产也不罚款就没有费用。又若三次试制后无合格品,则罚款 200 元,即 $f_4(1) = 2$ 。

以 $C(x_k)$ 表示生产成本及装配费用,则由每次装配费 200 元,每件成本 100 元得

$$C(x_k) = \begin{cases} 2 + x_k, & \text{当 } x_k > 0 \text{ 时} \\ 0, & \text{当 } x_k = 0 \text{ 时} \end{cases}$$

由生产一件得合格品的概率为 0.4,得不合格品的概率为 0.6,所以生产 x_k 件均不合格的概率应为 0.6^{x_k} ,至少有一件合格品的概率为 $(1 - 0.6^{x_k})$,这里 $x_k = 0, 1, 2, \dots$ 。于是递推关

系为

$$\begin{aligned} f_k(1) &= \min_{x_k \in D_k(1)} [C(x_k) + (1 - 0.6^k)f_{k+1}(0) + 0.6^k f_{k+1}(1)] \\ &= \min_{x_k \in D_k(1)} [C(x_k) + 0.6^k f_{k+1}(1)] \end{aligned}$$

这里 $D_k(1) = \{0, 1, 2, \dots\}$ 。于是当 $k=3$ 时,有

$$f_3(1) = \min_{x_3 \in D_3(1)} [C(x_3) + 0.6^3 \times 2]$$

对 x_3 的不同取值计算后得如表 9.28 的结果。

表 9.28

u_3	x_3	$C(x_3) + 20 \times 0.6^3$						$f_3(u_3)$	x_3
		0	1	2	3	4	5		
0	0	0	—	—	—	—	—	0	0
1	1	20	15	11.2	9.32	8.59	8.56	8.56	5

表中最后二列表示不同的 u_3 所对应的最优期望值及最优决策。当 $k=2$ 时,有

$$f_2(1) = \min_{x_2 \in D_2(1)} [C(x_2) + 8.56 \times 0.6^2]$$

计算后可得表 9.29 的结果。当 $k=1$ 时, $u_1=1$, 同样有

$$f_1(1) = \min_{x_1 \in D_1(1)} [C(x_1) + 6.85 \times 0.6^1]$$

表 9.29

u_2	x_2	$C(x_2) + 8.56 \times 0.6^2$					$f_2(u_2)$	x_2
		0	1	2	3	4		
0	0	0	—	—	—	—	0	0
1	1	8.56	8.14	7.08	6.85	7.11	6.85	3

计算后可得表 9.30 的结果。

表 9.30

u_1	x_1	$C(x_1) + 6.85 \times 0.6^1$				$f_1(u_1)$	x_1
		0	1	3	4		
1	1	6.85	7.11	6.46	6.48	6.46	2

以上三个表中,对 x_k 取较大数值时的 $C(x_k) + 0.6^k f_{k+1}(1)$ 之值没有列出来,但可以证明,以后的数值随 x_k 的增大而增大,是单调上升的。

至此,求得最优策略是:第一次生产 2 个;如果都不合格,则第二次生产 3 个;如果再都不合格,则第三次生产 5 个。这样能使期望费用最小,其期望费用为 646 元(近似值)。

习 题

9.1 设某工厂自国外进口一部精密机器,由机器制造厂至出口港有三个港口可供选择,而进口港又有三个可供选择,进口后可经由两个城市到达目的地,其间的运输成本如图 9.4 中所标的数字,试求运费最低的路线?

9.2 有一部货车每天沿着公路给四个零售店卸下 6 箱货物,如果各零售店出售该货物所得利润如表 9.31 所示,试求在各零售店卸下几箱货物,能使获得总利润最大? 其值是多少?

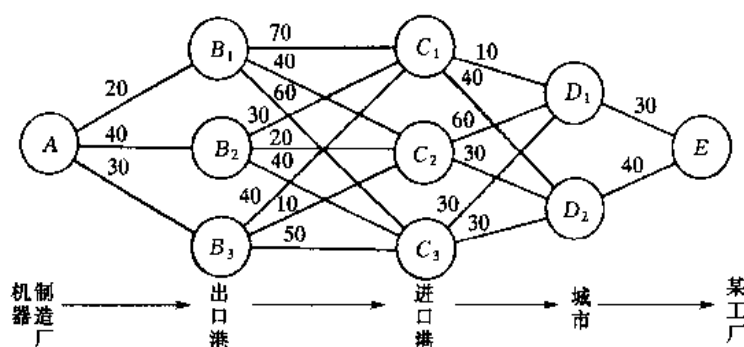


图 9.4

表 9.31

利润 箱数	零售店	1	2	3	4
0		0	0	0	0
1		4	2	3	4
2		6	4	5	5
3		7	6	7	6
4		7	8	8	6
5		7	9	8	6
6		7	10	8	6

9.3 某公司有三个工厂,它们都可以考虑改造扩建。每个工厂都有若干种方案可供选择,各种方案的投资及所能取得的收益如表 9.32 所示(单位:千万元)。现公司有资金 5 千万元,问应如何分配投资使公司的总收益最大?

表 9.32

m_i (方案)	工厂 $i=1$		$i=2$		$i=3$	
	C(投资)	R(收益)	C	R	C	R
1	0	0	0	0	0	0
2	1	5	2	8	1	3
3	2	6	3	9	—	—
4	—	—	4	12	—	—

(注:表中“—”表示无此方案)

9.4 某工厂在一年进行了 A、B、C 三种新产品试制,由于资金不足,估计在年内这三种新产品研制不成功的概率分别为 0.40、0.60、0.80,因而都研制不成功的概率为 $0.40 \times 0.60 \times 0.80 = 0.192$ 。为了促进三种新产品的研制,决定增拨 2 万元的研制费,并要资金集中使用,以万元为单位进行分配。其增拨研制费与新产品不成功的概率如表 9.33 所示。试问如何分配费用,使这三种新产品都研制不成功的概率为最小。

表 9.33

新产品 研制费 S	不 成 功 概 率		
	A	B	C
0	0.40	0.60	0.80
1	0.20	0.40	0.50
2	0.15	0.20	0.30

9.5 根据国家需要某工厂交货任务如表 9.34 所示。表中数字为月底的交货量。该厂的生产能力为每月 400 件,该厂仓库的存货能力为 300 件,已知每 100 件货物的生产费为 10 000 元,在进行生产的月份,工厂要支出经常费 4 000 元,仓库保管费为每百件货物每月 1 000 元。假定开始时及 6 月底交货后无存货。试问应在每个月各生产多少件物品才能既满足交货任务又使总费用最少?

9.34

月份	1	2	3	4	5	6
货物量(百件)	1	2	5	3	2	1

9.6 某人外出旅游,需将五件物品装入包裹,但包裹质量有限制,总质量不超过 13 千克。物品质量及其价值的关系如表 9.35 所示。试问如何装这些物品,使整个包裹价值最大?

表 9.35

物品	质量(kg)	价值(元)
A	7	9
B	5	4
C	4	3
D	3	2
E	1	0.5

9.7 设某企业在今后 5 年内需使用一台机器,该种机器的年收入、年运行费及每年年初一次性更新费用随机器的役龄变化如表 9.36 所示。该企业现有一台役龄为 1 年的旧机器,试制定最优更新计划,使在 5 年内的总收入达到最大(不计 5 年期末时机器的残值)。

表 9.36

机 龄	0	1	2	3	4	5
年收入	20	19	18	16	14	10
运行费	4	4	6	6	9	10
更新费	25	27	30	32	35	36

9.8 求下列问题的最优解

$$(1) \max z = 10x_1 + 22x_2 + 17x_3$$

$$\begin{cases} 2x_1 + 4x_2 + 3x_3 \leq 20 \\ x_i \geq 0 \text{ 且为整数 } (i=1,2,3) \end{cases}$$

$$(2) \max z = x_1 x_2 x_3 x_4$$

$$\begin{cases} 2x_1 + 3x_2 + x_3 + 2x_4 = 11 \\ x_i \geq 0 \text{ 且为整数 } (i=1,2,3,4) \end{cases}$$

第 10 章 多目标决策

10.1 特尔非法

特尔非法又称老手法,是美国兰德公司于 1964 年首先用于决策领域的。特尔非法是一种重要的多目标决策方法,它突破了传统的数量分析限制,为更合理地制定决策开阔了思路。它的主要优点是简明直观,避免了专家会议的许多弊病。例如,第 4 章讲述的线性加权函数的构造,关键就在于确定各目标的权重 w_j 。实践中常使用特尔非法确定权重。本节就以此为例,介绍特尔非法。

特尔非法是请一批有经验的专家(老手)对如何确定各目标权重发表意见,然后用统计平均方法估算出各目标的权重值。在此例中,首先应对各个目标值 f_j 标准化,即对每一目标事先给定一个标准值 f_{jH} ,然后将所有目标除以相应的标准值,即得无量纲化的目标值,即

$$\bar{f}_j = \frac{f_j}{f_{jH}} \quad (j=1,2,\cdots,p)$$

这种方法的具体步骤如下。

①把较为详尽的背景资料发送给选定的 n 位专家,请专家们分别各自独立估计各目标的权重,结果列入表 10.1 中。

表 10.1

专家数	目 标				
	权重估计值	f_1	f_2	...	f_p
1		w_{11}	w_{12}	w_{1p}
2		w_{21}	w_{22}	w_{2p}
\vdots	
n		w_{n1}	w_{n2}	w_{np}

②计算各目标权重的样本平均值,即

$$\bar{M}(w_j) = \frac{1}{n} \sum_{i=1}^n w_{ij} \quad (j=1,2,\cdots,p)$$

每一位专家对各目标权重估计值与平均估计值的偏差

$$\Delta_{ij} = w_{ij} - \bar{M}(w_j)$$

③进一步分析 $\bar{M}(w_j)$ 是否合理,特别让估计值偏差 Δ_{ij} 较大的专家充分发表意见,消除估计中的一些误差。

④附上进一步的补充资料后,请各专家重新对各目标权重值做出估计值 w_{ij} ,再一次计算平均估计值及方差

$$\tilde{M}(w_j) = \frac{1}{n} \sum_{i=1}^n w_{ij} \quad (j=1,2,\cdots,p)$$

$$\tilde{D}(w_j) = \frac{1}{n-1} \sum_{i=1}^n [w_{ij} - \tilde{M}(w_j)]^2$$

重复上述步骤,经过几次反复后,直至第 k 步估计方差小于或等于预先给定的标准 ε ($\varepsilon > 0$)。

⑤将第 k 步得到的平均估计值 $\tilde{M}(w_j)$ 及 $\tilde{D}(w_j)$ 再送交各位专家,请他们作最后的判断,给出各目标权重的最终估计值 w'_{ij} ($i=1,2,\dots,n, j=1,2,\dots,p$)。同时还要请每位专家标出各自对所给估计值的“信任度” l_{ij} ($i=1,2,\dots,n, j=1,2,\dots,p, l_{ij}$ 表示第 i 位专家就第 j 个目标权重估计值给出的“信任度”)。信任度表示专家对自己所做估计的把握程度,且规定信任度取值于 $[0,1]$,当专家有绝对把握时, $l_{ij}=1$;当专家毫无把握时, $l_{ij}=0$;除去上述两种极端情形外, $0 < l_{ij} < 1$ 。专家的信任度是一个心理指标,它取决于专家对资料信息的占有程度,取决于专家的经验与信念。当一个专家不占有充分的信息或资料时,是不大敢随意地作出某种判断的。可见,在特尔菲法中引入“信任度”,可以在一定程度上降低专家评判的主观效应,从而在一定程度上保证了权重确定过程的客观性。

⑥确定最终的目标函数权重估计值。令

$$M_{\lambda}^{(j)} = \{i: l_{ij} \geq \lambda, i=1,2,\dots,n\}$$

其中 λ 是预先给定的标准,且 $0 < \lambda < 1$ 。所以,第 j 个目标之权重的最终估计值

$$\bar{w}_j = \frac{1}{|M_{\lambda}^{(j)}|} \sum_{i \in M_{\lambda}^{(j)}} w'_{ij}$$

其中 $|M_{\lambda}^{(j)}|$ 表示集合 $M_{\lambda}^{(j)}$ 中元素的个数。

这种方法实质是先以 λ 为尺子,将信任度达不到 λ 的估计值全部删除,以余下估计值的平均值作为权重的最终估计值,因此,该方法有一定的合理性。

当然,各目标权重之最终估计值的确定也可采用其他方法。如果各位专家有平等的学术地位,不妨取其最后估计值的算术平均值作为最终权重估计值,否则,可取其加权算术平均值。限于篇幅,对此不再赘述。

最后,可构造线性加权评价函数为

$$U[F(\mathbf{X})] = \sum_{j=1}^p \bar{w}_j f_j(\mathbf{X})$$

10.2 层次分析法

层次分析法 (Analytic Hierarchy Process, 简称 AHP) 是由美国匹兹堡大学教授 T. L. Saaty 在 20 世纪 70 年代中期提出的。它的基本思想是把一个复杂的问题分解为各个组成因素,并将这些因素按支配关系分组,从而形成一个有序的递阶层次结构。通过两两比较的方式确定层次中诸因素的相对重要性,然后综合人的判断以确定决策诸因素相对重要性的总排序。层次分析法的出现给决策者解决那些难以定量描述的决策问题带来了极大方便,从而使它的应用涉及广泛的科学和实际领域。

10.2.1 简单情形——单层次模型

1. 单层次模型结构

如图 10.1 所示,此模型由一个目标 C 及隶属于它的 n 个评价元素 A_1, \dots, A_n 和决策者

组成。由决策者在这个目标意义下对这 n 个元素进行评价,对它们进行优劣排序并做出相对重要性的权衡,但由于决策者能力的限制很难一下子作出这种判断,而仅仅对两个元素进行优劣程度比较则是完全可能的。

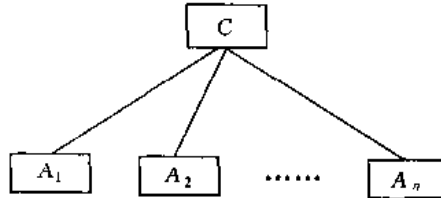


图 10.1

2. 思想

AHP 方法的基本思想就是将决策者对这 n 个元素优劣的整体判断转变为对这 n 个元素的两两比较,然后再转为对这 n 个元素的整体优劣排序判断及确定各元素的权重。

3. 计算步骤

(1) 构造两两比较判断矩阵

在单层次结构模型中,假定目标元素为 C_K ,同与之相连的有关元素 A_1, A_2, \dots, A_n 有支配关系。假定以上一层次某目标元素 C_K 作为准则,通过向决策者询问在原则 C_K 下元素 A_i 对元素 A_j 的优劣比较,构造一判断矩阵,其形式为:

C_K	A_1	A_2	A_n
A_1	a_{11}	a_{12}	a_{1n}
A_2	a_{21}	a_{22}	a_{2n}
\vdots	\vdots	\vdots	\vdots	\vdots
A_n	a_{n1}	a_{n2}	a_{nn}

其中 a_{ij} 表示对于 C_K 来说, A_i 对 A_j 相对重要性的数值体现,通常 a_{ij} 可取 1、2、.....、9 以及它们的倒数作为标度,含义如表 10.2 所示。

表 10.2 标度的含义

1	表示两个元素相比,具有同样重要性
3	表示两个元素相比,一个元素比另一个元素稍微重要
5	表示两个元素相比,一个元素比另一个元素明显重要
7	表示两个元素相比,一个元素比另一个元素强烈重要
9	表示两个元素相比,一个元素比另一个元素极端重要

2、4、6、8 为上述相邻判断的中值。

判断矩阵中的元素具有下述性质:① $a_{ij} > 0$; ② $a_{ij} = \frac{1}{a_{ji}}$; ③ $a_{ii} = 1$ 。

(2) 计算单一准则下元素的相对重要性(层次单排序)

这一步要根据判断矩阵计算对于目标元素而言各元素的相对重要性次序的权值。计算判断矩阵 A 的最大特征根 λ_{\max} 和其对应的经归一化后的特征向量 $W = [w_1 \quad w_2 \quad \dots \quad w_n]^T$ 。即首先对于判断矩阵 A 求解最大特征根问题:

$$AW = \lambda_{\max} W$$

得特征向量 W 并将其归一化, 将归一化后所得到的特征向量 $W = [w_1 \ w_2 \ \cdots \ w_n]^T$ 作为本层次元素 A_1, A_2, \dots, A_n 对于目标元素 C_k 的排序权值。

计算 λ_{\max} 和 W 一般采用近似计算的方根法, 步骤如下:

①将判断矩阵 A 中元素按行相乘: 即 $\prod_{j=1}^n a_{ij} (i=1, 2, \dots, n)$;

②计算 $\bar{w}_i = \sqrt[n]{\prod_{j=1}^n a_{ij}}$;

③将 \bar{w}_i 归一化得 $w_i = \frac{\bar{w}_i}{\sum_{j=1}^n \bar{w}_j}$, $W = [w_1 \ w_2 \ \cdots \ w_n]^T$ 为所求特征向量;

④计算最大特征根 $\lambda_{\max} = \sum_{i=1}^n \frac{(AW)_i}{nw_i}$, 其中, $(AW)_i$ 表示向量 AW 的第 i 个元素。

(3) 单层次判断矩阵 A 的一致性检验

在单层次判断矩阵 A 中, 当 $a_{ij} = \frac{a_{jk}}{a_{ki}}$ 时, 称判断矩阵为一致性矩阵。由于客观事物的复杂性和人们的偏爱不同, 判断矩阵很难有严格的一致性, 但应该要求有大致的一致性。因此, 在得到 λ_{\max} 后, 还需对判断矩阵的一致性进行检验。

进行一致性检验的步骤如下:

①计算一致性指标

$$C.I. = \frac{\lambda_{\max} - n}{n - 1}$$

式中 n 为判断矩阵的阶数。

②计算平均随机一致性指标 $R.I.$ 是多次重复进行随机判断矩阵特征值的计算后取算术平均数得到的, 表 10.3 给出 1~15 维矩阵重复计算 1,000 次的平均随机一致性指标值;

表 10.3

维数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$R.I.$	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.52	1.54	1.56	1.58	1.59

③计算一致性比例

$$C.R. = \frac{C.I.}{R.I.}$$

当 $C.R. < 0.1$ 时, 一般认为判断矩阵的一致性是可以接受的, 否则应修改矩阵使之符合要求。

4. 原理

将判断矩阵 A 的最大特征根 λ_{\max} 的特征向量经归一化后得 $W = [w_1 \ w_2 \ \cdots \ w_n]^T$, 将 W 作为本层次元素 A_1, A_2, \dots, A_n 的排序权重。对于目标元素 C_k 的排序权值的原理如下:

假定 n 个物体归一化后的重量分别为 w_1, w_2, \dots, w_n , 它们之间两两比较的相对重量可用矩阵表示为

$$A = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix}$$

若用重量向量 $W = [w_1 \ w_2 \ \cdots \ w_n]^T$ 右乘 A , 得

$$AW = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = nW$$

即得 $AW = nW$ 。

矩阵 A 具有以下特点:

- ① $a_{ii} = 1 \quad (i = 1, \cdots, n)$;
- ② $a_{ij} = 1/a_{ji} \quad (i, j = 1, 2, \cdots, n)$;
- ③ $a_{ij} = a_{ik}/a_{jk} \quad (i, j = 1, 2, \cdots, n)$ 。

根据矩阵理论可以证明, 该矩阵一定存在惟一的不为零的最大特征值 λ_{\max} , 且 $\lambda_{\max} = n$ 。

若 W 为未知, 在给出矩阵 A 的情况下, 可以通过求其特征值 λ_{\max} 及相应特征向量求出正规化的特征向量 W 作为 n 个对象的权重。同样, 对于复杂的社会、经济、管理等问题, 也可以通过建立层次分析结构模型, 构造出相应的判断矩阵 A , 应用上述原理确定各种方案、措施、成果等相对于总目标的优劣性或重要性的权重, 以供决策、评价等。

10.2.2 一般情形——多层次模型

利用 AHP 求解一般的多层次结构问题的基本步骤如下。

1. 建立递阶层次结构

首先对问题所涉及的因素进行分类, 构造一个各因素之间相互联结的递阶层次结构模型。处于最上面的层次一般是问题的预定目标, 通常只有一个元素, 中间层的元素一般是准则、子准则, 最低层一般是决策方案。一个典型的层次结构如图 10.2 所示。

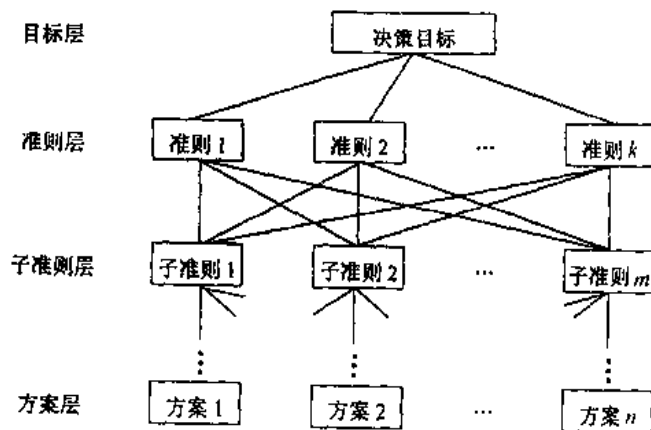


图 10.2

2. 构造两两比较判断矩阵

按前述方法构造各单层次的两两比较判断矩阵。

3. 计算单一准则下元素的相对重要性(单层次模型)

这一步要根据判断矩阵计算对上一层某元素而言本层次与之有联系的各元素的相对重要性次序的权值。即把各判断矩阵 A 视为单层次子模型,按上述单层次模型中的方法去求解特征根问题: $AW = \lambda_{\max} W$ 。

所得特征向量 W 经归一化后作为本层次元素 A_1, A_2, \dots, A_n 对于上一层元素的排序权值。

λ_{\max} 和 W 的计算方法前面已述及,在此不再重复。

4. 判断矩阵的一致性检验

在得到 λ_{\max} 和所对应的特征向量 $W = [w_1 \ w_2 \ \dots \ w_n]^T$ 后,还需要检验各判断矩阵的一致性。进行一致性检验的步骤与单层次模型完全相同。

5. 计算各层次上元素的组合权重(层次总排序)

层次总排序需要从上到下逐层进行。对于最高层,它的层次单排序即为总排序。

如果上一层所有元素 A_1, A_2, \dots, A_m 的组合权重已知,权值分别为 a_1, a_2, \dots, a_m , 与 A_i 相应的本层元素 B_1, B_2, \dots, B_n 的单排序结果为 $b_1^i, b_2^i, \dots, b_n^i (i=1, 2, \dots, m)$ 。若 B_j 与 A_i 无联系时, $b_j^i = 0$, 则本层次元素的组合权重可根据表 10.4 进行计算。显然有 $\sum_{j=1}^n b_j = 1$ 。

表 10.4

权重 层次 B	层次 A	A_1	A_2	A_m	B 层次元素组合权重
		a_1	a_2	a_m	
B_1		b_1^1	b_2^1	b_m^1	$b_1 = \sum_{i=1}^m a_i b_1^i$
B_2		b_1^2	b_2^2	b_m^2	$b_2 = \sum_{i=1}^m a_i b_2^i$
\vdots		\vdots	\vdots	\vdots	\vdots	\vdots
B_n		b_1^n	b_2^n	b_m^n	$b_n = \sum_{i=1}^m a_i b_n^i$

6. 评价层次总排序计算结果的一致性

为评价层次总排序计算结果的一致性,也需计算与层次单排序相类似的检验量。设 $C. I.$ 为层次总排序一致性指标; $R. I.$ 为层次总排序随机一致性指标。其计算公式为

$$C. I. = \sum_{i=1}^m a_i C. I. _i$$

$C. I. _i$ 为 A_i 相应的 B 层次中判断矩阵的一致性指标。

$$R. I. = \sum_{i=1}^m a_i R. I. _i$$

$R. I. _i$ 为 A_i 相对应的 B 层次中判断矩阵的随机一致性指标。

并取

$$C. R. = \frac{C. I. }{R. I. }$$

当 $C.R. < 0.10$, 认为层次总排序的结果具有满意的一致性。若不满足一致性条件, 需对判断矩阵进行调整。

AHP 的最终结果是得到相对于总的目标各决策方案的优先顺序权重, 并据此作出决策。

10.2.3 应用

例 10.1 某厂有一笔企业留成利润要决定如何使用, 根据各方意见, 提出的决策方案有发奖金、扩建集体福利设施、建图书馆、购买新设备。在决策时要考虑调动职工劳动积极性、提高职工技术文化水平、改善职工物质文化生活三方面, 据此构造各因素之间相互联结的层次结构模型如图 10.3 所示。

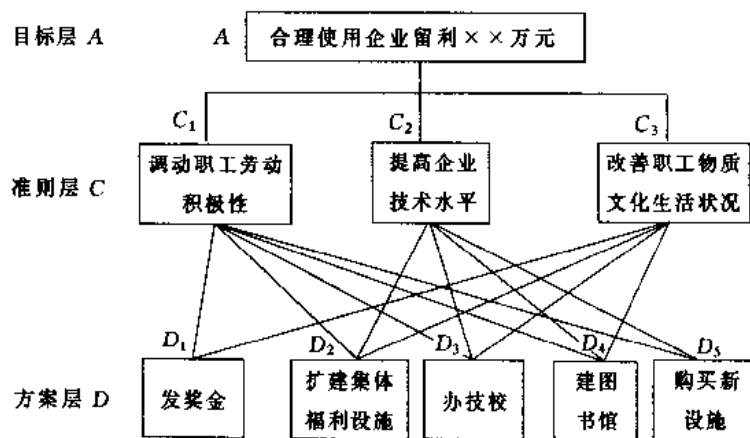


图 10.3

构造第二层相对于第一层的判断矩阵:

$A - C$	C_1	C_2	C_3
C_1	1	1/5	1/3
C_2	5	1	3
C_3	3	1/3	1

通过计算得判断矩阵的特征向量和特征值分别为

$$W = [0.105 \quad 0.637 \quad 0.258]^T$$

$$\lambda_{\max} = 3.308$$

对判断矩阵进行一致性检验, 即计算 $C.I.$ 和 $C.R.$:

$$C.I. = 0.019$$

$$C.R. = 0.033 < 0.1$$

说明判断矩阵的一致性可以接受。

同样可以构造第三层各元素相对于第二层元素的判断矩阵:

$C_1 - D$	D_1	D_2	D_3	D_4	D_5	$W = [0.491 \ 0.232 \ 0.092 \ 0.138 \ 0.046]^T$
D_1	1	2	3	4	7	$\lambda_{\max} = 5.126$
D_2	1/2	1	3	2	5	$C.I. = 0.032$
D_3	1/3	1/3	1	1/2	1	$C.R. = 0.028 < 0.1$
D_4	1/4	1/2	2	1	3	
D_5	1/7	1/5	1	1/3	1	
$C_2 - D$	D_2	D_3	D_4	D_5		$W = [0.055 \ 0.564 \ 0.118 \ 0.265]^T$
D_2	1	1/7	1/3	1/5		$\lambda_{\max} = 4.117$
D_3	7	1	5	2		$C.I. = 0.039$
D_4	3	1/5	1	1/3		$C.R. = 0.042 < 0.1$
D_5	5	1/2	3	1		
$C_3 - D$	D_1	D_2	D_3	D_4		$W = [0.406 \ 0.406 \ 0.094 \ 0.094]^T$
D_1	1	1	3	3		$\lambda_{\max} = 4$
D_2	1	1	3	3		$C.R. = 0$
D_3	1/3	1/3	1	1		
D_4	1/3	1/3	1	1		

第三层相对于第一层的权重通过第二层相对于第一层和第三层相对于第二层的权重组合而得到,计算结果如表 10.5 所示。

表 10.5

方案 \ 准则 权重	C_1	C_2	C_3	总权重
	0.105	0.637	0.258	
D_1	0.491	0	0.406	0.157
D_2	0.232	0.055	0.406	0.164
D_3	0.092	0.564	0.094	0.393
D_4	0.138	0.118	0.094	0.113
D_5	0.046	0.265	0	0.172

由此得到总权重向量

$$W = [0.157 \ 0.164 \ 0.393 \ 0.113 \ 0.172]^T$$

层次总排序一致性指标、层次总排序随机一致性指标为

$$C.I. = 0.028, \quad R.I. = 0.923, \quad C.R. = 0.03 < 0.10$$

计算结果表明,对于合理使用企业留成利润来说,办业余技校是首选方案。

10.3 数据包络分析法

数据包络分析(Data Envelopment Analysis, 简记 DEA)是著名运筹学家 A. Charnes 和 W. W. Cooper 等以相对效率概念为基础发展起来的一种新的效率评价方法。自 1978 年第一个 DEA 模型—— C^2R 模型(也称 CCR 模型)发表以来,有关的理论研究不断深入,应用领域日益广泛。事实证明,DEA 方法现已成为管理科学、系统工程、决策分析和评价技术等领域一种重要的分析工具和手段。

DEA 是使用数学规划模型比较决策单元间的相对效率,对决策单元做出评价的方法。一个决策单元(Decision Making Unit, 简记 DMU)在某种程度上是一种约定。它可以是学校、医院、法院、空军基地,也可以是银行或企业。其特点是:每个 DMU 都可以看做是相同的实体,亦即在某一视角下,各 DMU 具有相同的输入和输出。通过对输入输出数据的综合分析,DEA 可以得出每个 DMU 综合效率的数量指标,据此将各 DMU 定级排序,确定有效的(即相对效率最高的)DMU,并指出其他 DMU 非有效的原因和程度,给主管部门提供管理信息。DEA 还能判断各 DMU 的投入规模是否恰当,并给出了各 DMU 调整投入规模的正确方向和程度。

10.3.1 DEA 的基本模型—— C^2R 模型

1. C^2R 模型的基本形式

现在介绍 DEA 的最基本模型—— C^2R 模型。假设有 n 个决策单元,每个决策单元都有 m 种类型的“输入”,以及 s 种类型的“输出”,分别表示该单元“耗费的资源”和“工作的成效”,如图 10.4 所示。图中, x_{ij} 为第 j 个决策单元对第 i 种类型输入的投入量; y_{rj} 为第 j 个决策单元对第 r 种类型输出的产出量; v_i 为对第 i 种类型输入的一种度量(“权”); u_r 为对第 r 种类型输出的一种度量(“权”)。而且 $x_{ij} > 0, y_{rj} > 0, v_i \geq 0, u_r \geq 0 (i = 1, 2, \dots, m; r = 1, 2, \dots, s; j = 1, 2, \dots, n)$ 。 x_{ij} 及 y_{rj} 均为已知的数据,可以根据历史资料或预测得到, v_i 及 u_r 为“权”变量。

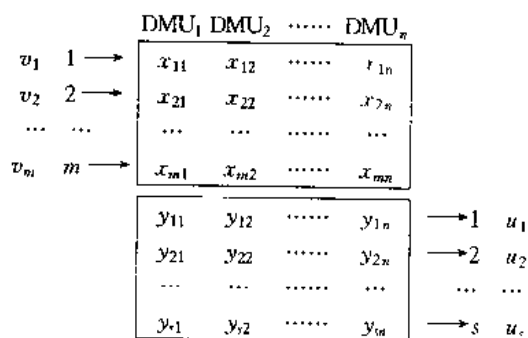


图 10.4

记 $X_j = [x_{1j} \quad \dots \quad x_{mj}]^T$, $Y_j = [y_{1j} \quad \dots \quad y_{sj}]^T (j = 1, \dots, n)$, 则可用 (X_j, Y_j) 表示第 j 个决策单元 DMU_j。

对应于权系数 $v = [v_1 \quad \dots \quad v_m]^T$, $u = [u_1 \quad \dots \quad u_s]^T$, 每个决策单元都有相应的效率评价指数

$$h_j = \frac{u^T Y_j}{v^T X_j} \quad (j=1, 2, \dots, n)$$

总可以适当选择权系数 v 和 u , 使其满足 $h_j \leq 1 (j=1, 2, \dots, n)$ 。

现在对第 j_0 个决策单元进行效率评价。简记 DMU_{j_0} 为 DMU_0 , $[X_{j_0} \ Y_{j_0}]$ 为 $[X_0 \ Y_0]$, h_{j_0} 为 $h_0, 1 \leq j_0 \leq n$ 。在各决策单元的效率评价指标均不超过 1 的条件下, 选择权系数 u 及 v , 使 h_0 最大。于是构成如下的最优化模型:

$$(\bar{P}) \begin{cases} \max & h_0 = \frac{u^T Y_0}{v^T X_0} = V_p \\ \text{s. t.} & h_j = \frac{u^T Y_j}{v^T X_j} \leq 1 \quad (j=1, 2, \dots, n) \\ & v \geq 0, u \geq 0 \end{cases}$$

这个原始规划模型是一个分式规划。利用 Charnes-Cooper 变换, 可以将 (\bar{P}) 化为一个等价的线性规划问题。令

$$t = 1/[v^T X_0], w = tv, \mu = tu$$

则原分式规划转化为

$$(P_1) \begin{cases} \max & \mu^T Y_0 = V_{P_1} \\ \text{s. t.} & w^T X_j - \mu^T Y_j \geq 0 \quad (j=1, 2, \dots, n) \\ & w^T X_0 = 1 \\ & w \geq 0, \mu \geq 0 \end{cases}$$

线性规划问题 (P_1) 的对偶规划问题为 (加入松弛向量 S^+ 及 S^- 以后)

$$(D_1) \begin{cases} \min & \theta = V_{D_1} \\ \text{s. t.} & \sum_{j=1}^n X_j \lambda_j + S^- = \theta X_0 \\ & \sum_{j=1}^n Y_j \lambda_j - S^+ = Y_0 \\ & \lambda_j \geq 0 \quad (j=1, 2, \dots, n); S^+ \geq 0; S^- \geq 0 \end{cases}$$

下面给出 C^2R 模型下 DEA 的有效定义。

定义 10.1 若线性规划问题 (P_1) 的最优解 w^0 及 μ^0 满足 $V_{P_1} = \mu^{0T} Y_0 = 1$, 则称 DMU_{j_0} 为弱 DEA 有效。

定义 10.2 若线性规划问题 (P_1) 存在某一最优解 w^0 与 μ^0 满足 $V_{P_1} = \mu^{0T} Y_0 = 1$, 并且 $w^0 > 0, \mu^0 > 0$, 则称 DMU_{j_0} 为 DEA 有效。

由定义易知, 若 DMU_{j_0} 为 DEA 有效, 那么它也是弱 DEA 有效。对于规划问题 (D_1) 有如下定理。

定理 10.1 ① DMU_{j_0} 为弱 DEA 有效的充要条件为规划问题 (D_1) 的最优值 $V_{D_1} = 1$; ② DMU_{j_0} 为 DEA 有效的充要条件为规划问题 (D_1) 的最优值 $V_{D_1} = 1$, 并且它的每个最优解 $\lambda^0 =$

$[\lambda_1^0 \cdots \lambda_n^0]^T, S^{-0}, S^{+0}, \theta^0$ 都满足 $S^{-0} = S^{+0} = 0$ 。

另外,可以证明,决策单元的 DEA 有效性等价于下述多目标规划问题的 Pareto 有效解

$$(V_P) \quad \begin{aligned} & \min F(X, Y) \\ & \text{s. t. } (X, Y) \in T_{CCR} \end{aligned}$$

其中: $X = [x_1 \cdots x_m]^T, Y = [y_1 \cdots y_s]^T; F(X, Y) = [X^T \quad -Y^T]^T; T_{CCR} = \{(X, Y) | X \geq \sum_{j=1}^n \lambda_j X_j, Y \leq \sum_{j=1}^n \lambda_j Y_j, \lambda_j \geq 0, j=1, \dots, n\}$ (称为 CCR 模型的经验生产可能集)。

在实际应用中,各输入量与输出量都带有一定的量纲。在不同的量纲下,输入量与输出量的数值不同,对此有下面定理。

定理 10.2 决策单元的最优效率评价指标 V_P 与输入量及输出量的量纲选取无关。

2. 具有非阿基米德无穷小量的 C^2R 模型

检验 DMU_{j_0} 的 DEA 有效性时,如果利用上面的线性规划问题 (P_1) ,需要判断是否存在最优解 w^0, u^0 满足:

$$w^0 > 0, u^0 > 0, V_{P_1} = \mu^T Y_0 = 1$$

如果利用对偶规划 (D_1) ,需要判断是否其所有最优解 $\lambda^0, S^{-0}, S^{+0}, \theta^0$ 都满足:

$$V_{D_1} = \theta^0 = 1, S^{-0} = 0, S^{+0} = 0$$

无论哪种方法都不方便。为此,Charnes 和 Cooper 等引进了非阿基米德无穷小量概念,构造了具有非阿基米德无穷小量的 C^2R 模型来判定 DMU_{j_0} 的有效性。

在广义实数域内,非阿基米德无穷小量 ϵ 是一个小于任何正数且大于零的“抽象数”。考虑带有非阿基米德无穷小量 ϵ 的 C^2R 模型 $(P_{1\epsilon})$:

$$(P_{1\epsilon}) \quad \begin{cases} \max & \mu^T Y_0 = V_{P_{1\epsilon}} \\ \text{s. t.} & w^T X_j - \mu^T Y_j \geq 0 \quad (j=1, 2, \dots, n) \\ & w^T X_0 = 1 \\ & w^T \geq \epsilon [e^-]^T, \mu^T \geq \epsilon [e^+]^T \end{cases}$$

$(P_{1\epsilon})$ 的对偶规划问题为

$$(D_{1\epsilon}) \quad \begin{cases} \min & [\theta - \epsilon([e^-]^T S^- + [e^+]^T S^+)] = V_{D_{1\epsilon}} \\ \text{s. t.} & \sum_{j=1}^n X_j \lambda_j + S^- = \theta X_0 \\ & \sum_{j=1}^n Y_j \lambda_j - S^+ = Y_0 \\ & \lambda_j \geq 0, j=1, 2, \dots, n; S^+ \geq 0; S^- \geq 0 \end{cases}$$

其中 $[e^-]^T = [1 \ 1 \ \cdots \ 1] \in E_m, [e^+]^T = [1 \ 1 \ \cdots \ 1] \in E_s$ 。

利用此模型,可以一次判断出 DMU_{j_0} 是 DEA 有效,还是仅为弱 DEA 有效,或者是非 DEA 有效。实际上有以下定理。

定理 10.3 设 ϵ 为非阿基米德无穷小量,并且规划问题 $(D_{1\epsilon})$ 的最优解为 $\lambda^0, S^{-0}, S^{+0}$,

θ^0 , 则有:

①若 $\theta^0 = 1$, 则 DMU_{j_0} 为弱 DEA 有效;

②若 $\theta^0 = 1$, 并且 $S^{-0} = 0, S^{+0} = 0$, 则 DMU_{j_0} 为 DEA 有效。

在实际应用中, 只要取 ϵ 足够小(例如取 $\epsilon = 10^{-6}$), 就可以使用单纯形方法求解规划问题 $(D_{1\epsilon})$ 。

例 10.2 考虑下述四个双输入、双输出的 DMU:

决策单元 评价指标	DMU ₁	DMU ₂	DMU ₃	DMU ₄
输入 1	2	3	4	3.5
输入 2	3	1	5.2	3.5
输出 1	1	2	2.5	4.3
输出 2	1.2	3.2	2.8	2.7

先对 DMU_1 进行分析, 为此建立如下面双输入的 CCR 模型:

$$\left\{ \begin{array}{l} \min \quad \theta - \epsilon \cdot (S_1^+ + S_2^+ + S_1^- + S_2^-) \\ \text{s. t.} \quad 2\lambda_1 + 3\lambda_2 + 4\lambda_3 + 3.5\lambda_4 + S_1^- = 2\theta \\ \quad \quad 3\lambda_1 + \lambda_2 + 5.2\lambda_3 + 3.5\lambda_4 + S_2^- = 3\theta \\ \quad \quad \lambda_1 + 2\lambda_2 + 2.5\lambda_3 + 4.3\lambda_4 - S_1^+ = 1 \\ \quad \quad 1.2\lambda_1 + 3.2\lambda_2 + 2.8\lambda_3 + 2.7\lambda_4 - S_2^+ = 1.2 \\ \quad \quad \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0, \lambda_4 \geq 0 \\ \quad \quad S_1^- \geq 0, S_2^- \geq 0, S_1^+ \geq 0, S_2^+ \geq 0 \end{array} \right.$$

令 $\epsilon = 10^{-6}$, 可求得最优解为

$$\theta^* = 0.6089, \lambda_2^* = 0.2943, \lambda_4^* = 0.0957, S_2^{-*} = 1.1974$$

因而根据定理 10.3 知, DMU_1 为非 DEA 有效。

类似地, 对于 DMU_2 、 DMU_3 和 DMU_4 可建立相应的规划模型, 并分别求得对应的最优解为

$$DMU_2: \theta^* = 1, \lambda_2^* = 1$$

$$DMU_3: \theta^* = 0.7258, \lambda_2^* = 0.6328, \lambda_4^* = 0.2871, S_2^{-*} = 2.1365$$

$$DMU_4: \theta^* = 1, \lambda_4^* = 1$$

于是根据定理 10.3, DMU_2 、 DMU_4 为 DEA 有效, 而 DMU_1 、 DMU_3 为非 DEA 有效。

上面从“产出不变、投入最少”的角度介绍了基于输入的 CCR 模型, 研究的是 DMU_0 的输入有效性。当然, 还可从“投入不变、产出最大”的角度出发研究 DMU_0 的输出有效性, 从而得到基于输出的 CCR 模型。在这两种模型中, 决策单元 DEA 有效性的定义是等价的。基于输出的 CCR 模型为

$$(P_{1\epsilon}) \left\{ \begin{array}{l} \min \quad w^T X_0 = V_{P_{1\epsilon}} \\ \text{s. t.} \quad w^T X_j - \mu^T Y_j \geq 0 (j = 1, 2, \dots, n) \\ \quad \quad \mu^T Y_0 = 1 \\ \quad \quad w^T \geq \epsilon [e^-]^T, \mu^T \geq \epsilon [e^-]^T \end{array} \right.$$

和

$$(D'_{1e}) \begin{cases} \max [\alpha + \varepsilon(e^{-T} S^- + e^{+T} S^+)] = V_{D'_{1e}} \\ \text{s. t. } \sum_{j=1}^n X_j \lambda_j + S^- = X_0 \\ \sum_{j=1}^n Y_j \lambda_j - S^+ = \alpha Y_0 \\ \lambda_j \geq 0; S^+ \geq 0; S^- \geq 0 \quad (j=1, 2, \dots, n) \end{cases}$$

10.3.2 DEA 有效性(C²R)的经济含义

1. 基本概念

(1) 生产可能集

DEA 方法一般都是基于一定的生产可能集概念讨论各 DMU 的生产有效性。

定义 10.3 称集 $T = \{(X, Y) | \text{产出 } Y \text{ 能用输入 } X \text{ 生产出来}\}$ 为所有可能的生产活动构成的生产可能集。

设某种生产活动的投入量 $X = [x_1 \ \cdots \ x_m]^T$, 产出量 $Y = [y_1 \ y_2 \ \cdots \ y_r]^T$, 于是可用 (X, Y) 表示该种生产活动。考虑 n 个决策单元, 对应的生产活动分别是 $(X_j, Y_j) (j=1, 2, \dots, n)$ 。通常称由 $(X_j, Y_j) (j=1, 2, \dots, n)$ 组成的集合

$$\hat{T} = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$$

为参考集。目的是根据上述参考集 \hat{T} 去估计生产可能集 T , 并确定哪些决策单元的生产活动是相对有效的。

在 C²R 模型中, 一般假设 T 满足以下公理。

1) 凸性 对任何 $(X, Y) \in T, (\hat{X}, \hat{Y}) \in T$, 对任何 $\lambda \in [0, 1]$, 都有:

$$\lambda(X, Y) + (1-\lambda)(\hat{X}, \hat{Y}) = (\lambda X + (1-\lambda)\hat{X}, \lambda Y + (1-\lambda)\hat{Y}) \in T$$

即若分别以 X 和 \hat{X} 的 λ 及 $1-\lambda$ 倍之和作为新的投入, 则可以得到原产出相同比例之和的新的产出。

2) 锥性 对任何 $(X, Y) \in T$ 和对任何 $k \geq 0$, 都有 $k(X, Y) = (kX, kY) \in T$ 。即若以投入量 X 的 k 倍进行输入, 那么输出量为原来产出的 k 倍是可能的。

3) 无效性 对任何 $(X, Y) \in T$, 都有 $(\hat{X}, Y) \in T$; 对任何 $\hat{X} \geq X, (X, \hat{Y}) \in T$, 对任何 $\hat{Y} \leq Y$ 。即在原来的基础上, 单方面增加投入或减少产出总是可能的生产活动, 也就是说, 浪费是存在的。

4) 最小性 生产可能集 T 是满足上述条件 1)~3) 的所有集合的交集。

显然观察到的经验生产活动 $(X_j, Y_j) \in T (j=1, 2, \dots, n)$, 在满足上述条件 1)~3) 的基础上, 对于已有的观测值 $(X_j, Y_j) (j=1, 2, \dots, n)$, 可得生产可能集

$$T_{\text{CCR}} = \{(X, Y) | k \cdot \sum_{j=1}^n \tau_j X_j < X, k \cdot \sum_{j=1}^n Y_j \tau_j \geq Y, \tau_j \geq 0, \sum_{j=1}^n \tau_j = 1, k > 0 \\ (j=1, 2, \dots, n)\}$$

若令 $k\tau_j = \lambda_j (j=1, 2, \dots, n)$, 上式变为

$$T_{CCR} = \{(X, Y) \mid \sum_{j=1}^n \lambda_j X_j \leq X, \sum_{j=1}^n Y_j \lambda_j \geq Y, \lambda_j \geq 0 (j=1, 2, \dots, n)\}$$

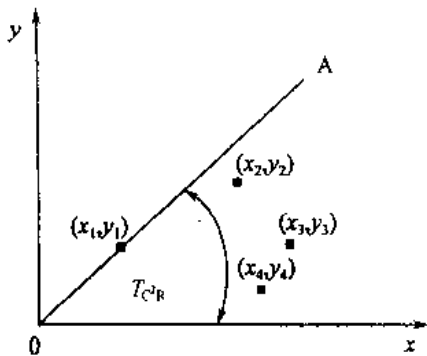


图 10.5

由于上述生产可能集是基于已有的观测值 (X_j, Y_j) ($j=1, 2, \dots, n$) 得到的, 习惯上称 T_{CCR} 为决策单元的经验生产可能集 (CCR), 它为一凸锥 (图 10.5)。

(2) 生产函数与规模收益

定义 10.4 设 $(X, Y) \in T$, 若不存在 $(X, Y') \in T$, 且 $Y \leq Y'$, 则称 (X, Y) 为“技术有效”生产活动。

定义 10.5 对生产可能集 T , 所有有效生产活动点 (X, Y) 构成的 R^{m+s} 空间的超曲面

$$Y = f(X)$$

称为生产函数。

显然, 生产函数表示在一定的技术条件下且生产处于最理想状态时, 投入量为 X 所能获得的最大产出量为 Y 。处于生产函数上的生产活动均为“技术有效”生产活动。易知生产函数中 Y 为关于 X 的增函数。

在经济生产活动分析中, 人们还常常使用规模收益的概念来反映产出 Y 对投入 X 的相对不减性程度, 为此, 给出如下定义。

定义 10.6 设 $(X, Y) \in T$, 令

$$\alpha(\beta) \triangleq \max \{ \alpha \mid (\beta X, \alpha Y) \in T, \beta \neq 1 \}$$

$$\rho \triangleq \lim_{\beta \rightarrow 1} \frac{\alpha(\beta) - 1}{\beta - 1}$$

若 $\rho > 1$, 称 (X, Y) 对应的 DMU 为规模收益递增; $\rho < 1$, 称为规模收益递减; $\rho = 1$, 称为规模收益不变。(这里假设上式极限总是存在的)

关于上述规模收益的概念, 可以从产出增量的相对百分比与对应投入增量的相对百分比的比值角度去理解。以单输入、单输出情形为例, 设其生产函数为图 10.6 中的 $PBAQ$ 折线。易知, 在 BP 线, $\rho = \frac{\Delta y}{y} / \frac{\Delta x}{x} > 1$, 对应的生产活动为规模收益递增; 在 BA 段, $\rho = \frac{\Delta y}{y} / \frac{\Delta x}{x} = 1$, 为规模收益不变; 而 AQ 段, $\rho = \frac{\Delta y}{y} / \frac{\Delta x}{x} < 1$, 为规模收益递减。

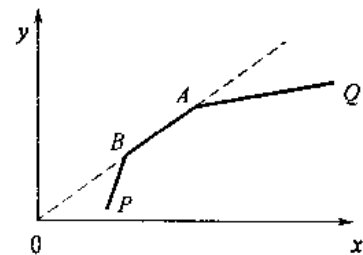


图 10.6

如果某一生产过程 (X_0, Y_0) 处于规模收益递增状态, 说明在 X_0 的基础上, 适当增大投入量, 可望最大可能产出有相对更高比例的增加, 因而 DMU_0 会有增加投入的积极性。反之, 从理论上说, DMU_0 将没有再增加投入的积极性。

如果对投入 X_0 , 当投入规模小于 X_0 时, 均为规模收益递增状态, 而当投入大于 X_0 时, 则相反。换句话说, 就投入规模而言, 无论大于或小于 X_0 都不是最好的, 称这样的 DMU_0 为规模收益不变或规模有效。

2. DEA 有效性 (CCR) 的经济含义

检验 DMU_{j_0} 的 DEA 有效性, 即是考虑规划问题:

$$(D_1) \begin{cases} \min \theta = V_{D_1} \\ \text{s. t. } \sum_{j=1}^n X_j \lambda_j \leq \theta X_0 \\ \sum_{j=1}^n Y_j \lambda_j \geq Y_0 \\ \lambda_j \geq 0 \quad (j=1, 2, \dots, n) \end{cases}$$

由于 (X_0, Y_0) 位于生产可能集 T_{CCR} 内, 由

$$T_{CCR} = \{(X, Y) \mid \sum_{j=1}^n \lambda_j X_j \leq X, \sum_{j=1}^n Y_j \lambda_j \geq Y, \lambda_j \geq 0 (j=1, 2, \dots, n)\}$$

可以看出, 规划问题 (D_1) 致力于在生产可能集 T 内, 保持产出 Y_0 不变, 同时将投入量 X_0 按同一比例尽量减少。如果 X_0 不能按同一比例减少, 即 (D_1) 的最优值 $V_0 = \theta^0 = 1$, 则表明 DMU_{j_0} 是有效的生产活动。

根据前面 DEA 有效的定义和生产活动技术有效的定义可知, DEA 有效 (C^2R) 的 DMU 必是技术有效 DMU。同时, C^2R 定义的生产可能集 T_{CCR} 为凸锥 (图 10.5), 可证 DEA 有效的 DMU_{j_0} 对应的生产点 (X_0, Y_0) 必位于此凸锥生产前沿面 OA 上, 而 OA 对应的各生产点均是规模有效生产点 (参见前面规模收益不变即规模有效的定义及图 10.6), 因而 DEA 有效的 DMU 必规模有效。

综上所述, DEA 有效 (C^2R) 等价于技术有效和规模有效。因此, 在 C^2R 模型中, 决策单元的 DEA 有效性是同时针对规模有效性和技术有效性而言的。

同时, 由 C^2R 的规划问题 (D_1) 的最优解可以进一步判断 DMU_{j_0} 的规模收益状况。设

(D_1) 具有惟一最优解 $\lambda^0, S^{-0}, S^{+0}, \theta^0$, 可以证明, 当 $\sum_{j=1}^n \lambda_j^0 / \theta^0 = 1$ 时, DMU_{j_0} 具有恰当的投入规模, 为规模收益不变; 当 $\sum_{j=1}^n \lambda_j^0 / \theta^0 < 1$ 时, DMU_{j_0} 为规模收益递增; 当 $\sum_{j=1}^n \lambda_j^0 / \theta^0 > 1$ 时, DMU_{j_0} 为规模收益递减。

例 10.3 如何科学地对普通中学进行评估, 这是近几十年一直在研究的问题。合理、准确的评估方法是一种积极的导向, 有助于端正教育思想, 引导正确的方向。特别是义务教育阶段, 积极推行综合素质教育, 避免片面追求升学率, 尤为需要建立一个科学、合理的评估方法与模型。教育评估和一般企业评估一样, 不能只从“产值”角度出发, 还应该充分考虑“效率”角度, 即应考虑其产出与投入的比值。下面以某市市区各普通初级中学为对象, 应用前面所介绍的 DEA 模型, 对其办学情况进行相对有效性评价。

考虑到该市各学校所处的经济、社会、文化环境, 所选决策单元为该市市区 20 所普通中学。评估模型所用输入输出指标体系如下。

输出: Y_1 毕业生人数, 按 90 届毕业生人数考虑; Y_2 毕业生平均成绩, 以 90 届毕业生毕业统考的入均成绩计算; Y_3 毕业生的身体素质, 以 1990 届毕业生的体育达标率计算。

输入: X_1 表示师资力量, $X_1 = ? z_1 + 1.5 z_2 + 1.2 z_3 + z_4$, 其中 z_1, z_2, z_3 分别为学校在编的特级、一级、二级教师人数, z_4 为其他人数; X_2 教育经费, 按 1990 年度下拨教育经费计算;

X_3 表示仪器设备、图书资料总额,按截止到 1990 年 8 月普通初中所拥有仪器设备、图书资料总额计算。

以上各数据的统计与计算结果请见表 10.5。

表 10.5

序号	中学名称	X_1	X_2	X_3	Y_1	Y_2	Y_3
1	市二中	92	25.2	2.82	354	412	1
2	市四中	80	24	2.87	227	368	1
3	市六中	72	18	3.57	120	378	1
4	市十中	68	14	2.75	181	376	1
5	市九中	90	15.4	2.38	174	341	1
6	市十一中	98	22.8	12	220	300	1
7	市十二中	98	21.95	10.3	223	323	1
8	北二中	45	2	1.4	139	441	1
9	北三中	34	14.5	2.22	91	409	0.87
10	山一中	66	8.5	2.54	174	356	1
11	山二中	57	6.3	1.83	140	360	1
12	山三中	57	13.8	7.47	158	296	0.98
13	山四中	71	12.8	1.7	137	345	0.91
14	山五中	61	12.7	2.2	145	342	1
15	耀化中学	61	22	3.75	129	362	0.85
16	秦铁中	69	4.5	3.8	118	345	1
17	秦附中	34	10	1.5	128	349	1
18	农技中	29	12.6	8.42	76	248	0.94
19	黄庄中学	43	11.54	3.42	43	288	1
20	山桥中学	46	15.2	5.65	167	409	1

利用 C^2R 模型,分别对上述各学校(决策单元)建立相应的线性规划模型,求得各中学的评价结果见表 10.6。

表 10.6

中学名称	对应 C^2R 模型最优解	评价结论
市二中	$\theta^* = 1, \lambda_1^* = 1$	DEA 有效,规模收益不变
市四中	$\theta^* = 0.7426, \lambda_1^* = 0.4381, \lambda_{17}^* = 0.5619, S_2^* = 1.1637, S_3^* = 0.053, S_2^{+*} = 8.5973$	非 DEA 有效,规模收益不变
市六中	$\theta^* = 0.5078, \lambda_8^* = 0.188, \lambda_9^* = 0.111, \lambda_{17}^* = 0.7154, S_3^* = 0.2299, S_1^{+*} = 7.8075$	非 DEA 有效,规模收益递减
市十中	$\theta^* = 0.7435, \lambda_1^* = 0.2169, \lambda_8^* = 0.3611, \lambda_{17}^* = 0.4219, S_3^* = 0.2943, S_2^{+*} = 19.8915$	非 DEA 有效,规模收益不变
市九中	$\theta^* = 0.6854, \lambda_1^* = 0.1628, \lambda_8^* = 0.8372, S_1^* = 9.0315, S_2^* = 4.7778, S_2^{+*} = 95.2791$	非 DEA 有效,规模收益不变

续表

中学名称	对应 C_2R 模型最优解	评价结论
市十一中	$\theta^* = 0.609\ 7, \lambda_1^* = 0.394\ 3, \lambda_8^* = 0.261\ 6, \lambda_{17}^* = 0.344,$ $S_3^* = 5.321\ 9, S_2^* = 97.913\ 7$	非 DEA 有效, 规模收益不变
市十二中	$\theta^* = 0.621\ 9, \lambda_1^* = 0.405\ 1, \lambda_8^* = 0.313\ 4, \lambda_{17}^* = 0.281\ 5,$ $S_3^* = 4.401\ 9, S_2^* = 80.356\ 5$	非 DEA 有效, 规模收益不变
北二中	$\theta^* = 1, \lambda_8^* = 1$	DEA 有效, 规模收益不变
北三中	$\theta^* = 1, \lambda_8^* = 1$	DEA 有效, 规模收益不变
山一中	$\theta^* = 0.785, \lambda_1^* = 0.167\ 8, \lambda_8^* = 0.734\ 7, \lambda_{17}^* = 0.097\ 5, S_3^* =$ $0.346, S_2^* = 71.160\ 5$	非 DEA 有效, 规模收益不变
山二中	$\theta^* = 0.778\ 1, \lambda_1^* = 0.009\ 8, \lambda_8^* = 0.889\ 6, \lambda_{17}^* = 0.100\ 6,$ $S_2^* = 1.869\ 7, S_2^* = 71.456\ 3$	非 DEA 有效, 规模收益不变
山三中	$\theta^* = 0.756\ 6, \lambda_1^* = 0.135\ 4, \lambda_8^* = 0.177\ 2, \lambda_{17}^* = 0.667\ 3,$ $S_3^* = 4.020\ 6, S_2^* = 70.858\ 6$	非 DEA 有效, 规模收益递增
山四中	$\theta^* = 0.790\ 2, \lambda_1^* = 0.048\ 9, \lambda_8^* = 0.861\ 1, S_1^* = 12.859\ 8, S_2^* =$ $7.161, S_2^* = 54.892\ 4$	非 DEA 有效, 规模收益递增
山五中	$\theta^* = 0.691\ 9, \lambda_1^* = 0.052\ 4, \lambda_8^* = 0.469\ 8, \lambda_{17}^* = 0.477\ 8,$ $S_2^* = 1.749\ 5, S_2^* = 53.520\ 6$	非 DEA 有效, 规模收益不变
耀化中学	$\theta^* = 0.572, \lambda_8^* = 0.063\ 9, \lambda_{17}^* = 0.962\ 4, S_2^* = 2.034\ 4,$ $S_3^* = 0.559\ 7, S_3^* = 0.168$	非 DEA 有效, 规模收益递减
秦铁中	$\theta^* = 0.635\ 1, \lambda_8^* = 0.892\ 8, \lambda_{17}^* = 0.107\ 2, S_3^* = 1.002\ 6, S_1^* =$ $19.820\ 4, S_2^* = 861.134\ 7$	非 DEA 有效, 规模收益不变
秦附中	$\theta^* = 1, \lambda_{17}^* = 1$	DEA 有效, 规模收益不变
农枝中	$\theta^* = 1, \lambda_{18}^* = 1$	DEA 有效, 规模收益不变
黄庄中学	$\theta^* = 0.811\ 1, \lambda_8^* = 0.079\ 9, \lambda_{17}^* = 0.920\ 1, S_3^* = 1.282\ 1, S_1^* =$ $85.879\ 2, S_2^* = 68.353\ 1$	非 DEA 有效, 规模收益不变
山桥中学	$\theta^* = 0.960\ 6, \lambda_1^* = 0.083\ 8, \lambda_{17}^* = 1.073, S_2^* = 1.760\ 5,$ $S_3^* = 3.581\ 8, S_3^* = 0.156\ 8$	非 DEA 有效, 规模收益递减

由表 10.6 知, DEA 有效的中学有市二中、北二中、北三中、秦附中和农枝中, 其他中学均为非 DEA 有效。进一步地, 可对各非有效中学进行 DEA 投影分析, 见表 10.7。

表 10.7

决策单元	评价指标		原始指标数据	DEA 相对有效面“投影”面
市四中	投入指标	X_2	24	16.658 7
		X_3	2.87	2.078 3
	产出指标	Y_2	368	376.597 3
市六中	投入指标	X_3	3.57	1.582 9
	产出指标	Y_1	120	127.807 5
市十中	投入指标	X_3	2.75	1.750 3
	产出指标	Y_2	376	395.891 5

续表

决策单元	评价指标		原始指标数据	DEA 相对有效面“投影”面
市九中	投入指标	X_1	90	52.654 5
		X_2	15.4	5.777 4
	产出指标	Y_2	341	436.279 1
市十一中	投入指标	X_3	12	1.994 5
	产出指标	Y_2	300	397.913 7
市十二中	投入指标	X_3	10.3	2.003 7
	产出指标	Y_2	323	403.356 5
山一中	投入指标	X_3	2.54	1.647 9
	产出指标	Y_2	356	427.160 5
山二中	投入指标	X_2	6.3	3.032 3
	产出指标	Y_2	360	431.456 3
山三中	投入指标	X_3	7.47	1.631 2
	产出指标	Y_2	296	366.858 6
山四中	投入指标	X_1	71	43.244 4
		X_2	12.8	2.953 6
	产出指标	Y_2	345	399.892 4
山五中	投入指标	X_2	12.7	7.037 6
	产出指标	Y_2	342	395.520 6
耀化中学	投入指标	X_2	22	10.549 6
		X_3	3.75	1.585 3
	产出指标	Y_3	0.85	1.018
秦铁中	投入指标	X_3	3.8	1.410 8
	产出指标	Y_1	118	137.820 4
		Y_2	345	431.134 7
黄庄中学	投入指标	X_3	3.42	1.491 9
	产出指标	Y_1	43	128.879 2
		Y_2	288	356.353 1
山桥中学	投入指标	X_2	15.2	12.840 6
		X_3	5.65	1.845 6
	产出指标	Y_3	1	1.156 8

普通中学相对有效性的评价结果既反映了学校内部的教学、管理等问题,同时也反映了上层管理部门在进行办学经费配置和其他决策上的问题。因此对于非 DEA 有效的学校,一方面在现有的办学条件上要努力工作、提高教学质量、加强教学管理;另一方面,上层管理部门应对该学校的办学规模等进行适当的调整,使该学校成为 DEA 有效。

例 10.4 (某县工业产业经济有效性评价实例)某县的工业伴随改革开放的大潮得到了

较快发展,在全县国民经济发展中占有主要地位。为了进一步优化调整产业结构,依据该县工业的有关统计数据,应用 DEA 方法的 CCR 模型,对该县的工业经济进行了相对有效性评价。

该县工业的产业结构及各产业的序号如图 10.7 所示。



图 10.7

其中,“其他”产业是指该县一些规模过小的产业,如木材加工与家具制造、塑料制品、金属制品业等的总称。

投入、产出指标及其代表符号见图 10.8。

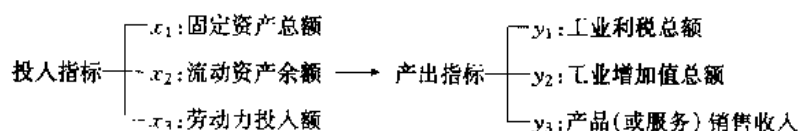


图 10.8

根据 1996 年该县国民经济统计资料汇编等资料,得到的该县 1995 年的投入产出数据如表 10.8 所示。

表 10.8

序号	要素 行业	投入要素			产出要素		
		固定资产 总额	流动资产 总额	劳动力投 入额	工业利税 总额	工业增加 值总额	产品销售 收入
1	建材	11 046	15 162	4 050	4 491	8 483	35 871
2	食品	9 290	12 709	1 900	2 218	5 714	28 940
3	服装	10 343	17 880	3 435	2 920	8 928	32 033
4	机械	8 447	16 615	2 008	2 260	4 929	23 655
5	纺织	7 839	10 011	1 325	166	2 241	12 393
6	冶金	1 463	1 826	384	754	1 875	7 346
7	化学	6 252	7 213	783	629	2 643	11 746
8	造纸	2 391	4 527	749	889	1 677	9 486
9	电子	1 670	1 154	345	807	1 738	6 326
10	工艺品	821	2 420	332	249	749	2 695
11	其他	4 486	8 477	872	1 531	2 635	14 049

采用 CCR 模型求解,得到评价各产业相对有效性的 θ 值及规模收益情况,如表 10.9 所示。

表 10.9

序号	产业	DEA 得分(θ 值)
1	建材	0.744(规模效益递减)
2	食品	0.934(规模效益递减)
3	服装	0.731(规模效益递减)

续表

序号	产业	DEA 得分(θ 值)
4	机械	0.915(规模效益递减)
5	纺织	0.565(规模效益递减)
6	冶金	1.000
7	化学	1.000
8	造纸	0.901(规模效益递减)
9	电子	0.716(规模效益递增)
10	工艺品	1.000
11	其他	1.000

由表可以看出,1995 年该县工业产业的相对有效性按 DEA 评价为有效的产业有冶金、化学、工艺品和其他,按 DEA 评价较差的产业有纺织和电子。这与该县的实际情况是相符的。

习 题

某公司欲确定下一年度广告宣传方式,宣传媒介有街头广告牌(C_1)、报纸(C_2)和电视(C_3)三种。公司选择宣传方式的标准有观众(读者)人数(B_1)、宣传效果(B_2)和广告费用(B_3)。已知各元素的相对重要度如下:

A	B_1	B_2	B_3	W_1^0	B_1	C_1	C_2	C_3	W_1^1
B_1	1	2	2	0.484	C_1	1	1/3	1/5	0.105
B_2		1	1/3	0.168	C_2		1	1/3	0.258
B_3			1	0.348	C_3			1	0.637
B_2	C_1	C_2	C_3	W_2^1	B_3	C_1	C_2	C_3	W_2^2
C_1	1	2	7	0.592	C_1	1	1/3	1/7	0.081
C_2		1	5	0.333	C_2		1	1/5	0.188
C_3			1	0.075	C_3			1	0.731

要求:(1)建立多级递阶结构;(2)排列三种宣传方式的优先顺序。

V 对策分析技术

第 11 章 二人有限零和对策

11.1 基本概念

11.1.1 对策现象与对策论

1. 对策和对策论

在日常生活中及各种领域内,经常可以看到一些充满着竞争、对抗、冲突的现象。对策论是研究上述现象的数学理论与方法。它是一种理论模型,其中包括参加者所掌握的全部信息及可能采取的行动等。

对策论把各式各样的冲突现象抽象成一种数学模型,然后给出分析这些问题的方法和解。应该说明的是,所谓解是指对策中的所有参加者都按最佳策略行动而得到的结果。对策论的研究中一般都假设:在对策中所有参加者都是“完全理智”的,在采取的策略上没有任何失误。

2. 对策理论的历史

在人类生活中到处都存在着“冲突”,所以人类在处理对策问题方面已有一个很长的历史。在我国春秋战国时期成书的“孙子兵法”就是运用对策知识于战争中的经验总结。围棋也是一种对策形式,它发明于我国殷代。自 19 世纪,西方人开始把对策知识应用于经济领域。

对策作为一种数学理论开始于 1944 年。纽曼(Neumann)和曼格斯特(Morgenstern)发表了题为《对策论及在经济领域中的应用》的著作之后,很快引起了数学家、社会学家和经济学家的极大兴趣。对策论之所以能在 40 年代得到迅速发展,主要原因是第二次世界大战中在军事问题上,人们总想找到一种对策模型,从而能得到正确的解。第二次世界大战结束后,人们又想到把对策理论应用到非军事的各个领域。特别是 20 世纪 90 年代以来对策理论在金融、管理和经济领域中得到了广泛应用。

3. 对策问题举例

(1) 中国的游戏——“剪刀、石头、布”

这在中国的孩子中广泛流行。当二个孩子在玩什么游戏之前决定前后顺序时,他们先作“猜手”。方法是:每个人有布(五指平伸)、石头(拳头)和剪子(食指和中指平伸,其余三指弯曲收紧)三种手势,玩时二个孩子独立地选择三种手势中的任何一种,然后同时展示给对方。决定胜负的规则是:布胜石头,石头胜剪子,剪子胜布。如果二个孩子同时伸出了相同的手势,被看成是一次平局,这需要继续猜,直至决出胜负为止。

(2) 罪犯二难推理

假设有二个人被怀疑有抢劫犯罪行为而被捕,并被分别关在二个牢房里,互相不许接触。这二个人确实犯了抢劫罪,可这二个人都清楚,如果他们都不坦白,那么就没有足够的证据指控他们犯罪,而只能因藏有枪支而被判 1 年徒刑;如果两个人都坦白了,则每个人都判 9 年徒刑;如果一个坦白,而另一个不坦白,则坦白的被无罪释放,而不坦白的则要判 10 年徒刑。这二个人应该怎样做呢?

(3) 齐王赛马

战国时期,齐国国王有一天提出要与大将军田忌赛马。田忌答应后,双方约定:

- ①每人从上中下三个等级中各出一匹马,共出三匹;
- ②一共比赛三次,每一次比赛各出一匹马;
- ③每匹被选中的马都得参加比赛,而且只能参加一次;
- ④每次比赛后输者要付给胜者一千金。

当时在三个不同等级中,齐王的马要比田忌的强些,看来田忌要输三千金了,但由于田忌采用了谋士孙臆的意见,最终反败为胜。谋士的主意是:

- ①每次比赛前先让齐王说出他要出哪匹马;
- ②让田忌用下马对齐王上马;
- ③用中马对齐王下马;
- ④用上马对齐王中马。

4. 对策的分类

对策可分为静态对策与动态对策两大类。而静态对策又可分为二人对策和多人对策二种,多人对策又可分为结盟对策和不结盟对策。此外,根据策略集的有限或无限、得失函数是否为零,又可分为种种类型的对策模型,如图 11.1 所示。

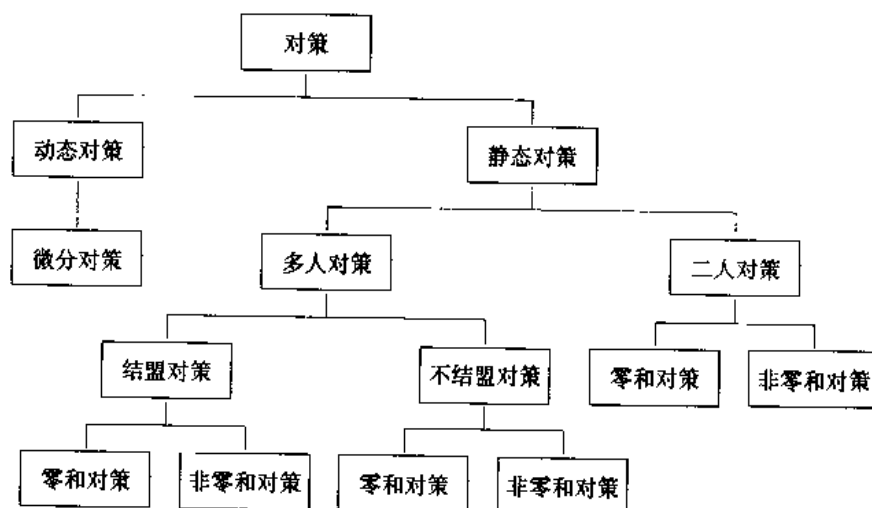


图 11.1

11.1.2 对策问题的组成

一般对策问题中有如下几个基本要素,即局中人、策略、局势、得失值。

1. 局中人(参加者)

对策中有决策权的参加者称为局中人。在齐王赛马的例子中,田忌是局中人,而孙臧不是。局中人除理解为个人外,还可理解为集体。有时大自然也可理解为局中人。为了便于研究问题,在对策中一些利益完全一致的参加者们可以结合成为一个联盟。他们齐心协力、相互配合、统一行动,称为结盟的一种形式。如桥牌游戏中,虽然有四个人参加,但只组合成两个局中人。

2. 策略与策略集合

一局对策中,把局中人的一个可行的行动方案称为他的一个策略,策略可以只含有一步行动方案。在比较复杂的对策中,可以是由自始至终指导行动的一系列步骤组成,如在齐王赛马的例子中,参加者双方的每个策略中都含有三个行动方案。通常一个局中人有好几个策略可供选择,把局中人的策略全体叫策略集合。用

$$S = \{s_1, s_2, \dots, s_m\}$$

表示局中人甲的策略集合, s_i 为局中人甲可选择的策略。同样用

$$D = \{d_1, d_2, \dots, d_n\}$$

表示局中人乙的策略集合, d_j 为局中人乙可选择的策略。

3. 局势

当每个局中人从各自策略集合中选择一策略而组成的策略组称为一个局势,用 (s_i, d_j) 表示。

4. 得失值

得失值是指局中人选定某局势后相应的收益值,又称支付。例如,当局中人甲选择 s_i 策略和局中人乙选 d_j 策略时,局中人甲的得失值可用 $R_{\text{甲}}(s_i, d_j)$ 表示,局中人乙的得失值可用 $R_{\text{乙}}(s_i, d_j)$ 表示,显然“得失值”是局势的函数。这个得失值既可以是一般意义下的收益或损失,也可以是局中人对局势的效用值。

11.2 二人有限零和对策的纯策略对策模型

二人有限零和对策的特点是:

①对策中只有两个局中人,双方的策略集

$$S = \{s_1, s_2, \dots, s_m\} \text{ 和 } D = \{d_1, d_2, \dots, d_n\}$$

均是有限集;

②在零和对策中,双方收益之和为零,甲的收益就是乙的损失,即 $R_{\text{甲}}(s_i, d_j) = -R_{\text{乙}}(s_i, d_j)$ 。

所以,在两人零和对策中,只需分析局中人甲的得失即可。通常可以用支付矩阵 $A = [a_{ij}]_{m \times n}$ 表示,其中 $a_{ij} = R_{\text{甲}}(s_i, d_j)$ 。因此,二人有限零和对策又称矩阵对策。

11.2.1 纯策略对策模型

人们称 $G = \{S, D, A\}$ 为一个纯策略对策模型,其中 $S = \{s_1, s_2, \dots, s_m\}$, $D = \{d_1, d_2, \dots, d_n\}$ 分别为局中人甲和乙的策略集合, $A = [a_{ij}]_{m \times n}$ 为局中人甲的支付矩阵。

11.2.2 纯策略对策的解

1. 纯策略分析

对策中,双方局中人都将确定自己的最佳策略,其实质也是一个决策过程。因此每个局中人同样首先得对自己的各个策略(行动方案)按某准则作一评价,只是在对策中要充分考虑对方的策略选择所造成的影响,实际是一种竞争性决策。

局中人甲认为当自己选择 s_i 策略时,对方必然选择对自己最不利的策略,即使自己得失值最小的策略,因此局中人甲对每个策略 s_i 的评价值为 $f(s_i)$ 。

$$s_i \xleftarrow{\text{评价}} f(s_i) = \min_j a_{ij}$$

因此局中人甲选择的策略模型为

$$s_i^* \leftarrow \max_i f(s_i) = \max_i \min_j a_{ij} = V_{\max}$$

同样,局中人乙认为自己选择策略 d_j 时,局中人甲会选择对乙最不利的策略,即使局中人甲收益最大的策略。因此,他对策略 d_j 的评价值 $g(d_j)$ 为

$$d_j \xleftarrow{\text{评价}} g(d_j) = \max_i a_{ij}$$

因此局中人乙确定策略的决策模型为

$$d_j^* \leftarrow \min_j g(d_j) = \min_j \max_i a_{ij} = V_{\min}$$

2. 纯策略对策模型的解

(1) 鞍点与解

对于一个对策 $G = \{S, D, A\}$ 。如果有

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = a_{i^*, j^*}$$

则称局势 (s_i^*, d_j^*) 为对策 G 的一个鞍点, s_i^* 为甲的最优纯策略, d_j^* 为乙的最优纯策略。 $V = a_{i^*, j^*}$ 为对策 G 之值, (s_i^*, d_j^*) 为对策的一个解,此时称对策 G 有解。

对于一个对策 G ,很容易证明下述定理。

定理 11.1 对于纯策略对策 $G = \{S, D, A\}$,总有

$$\min_j \max_i a_{ij} \geq \max_i \min_j a_{ij}$$

显然,对局中人甲而言,不管乙采用什么策略,他至少能得到的得失值为 $V_L = \max_i \min_j a_{ij}$;而对局中人乙而言,他能保证在任何情况下,甲的得失值不会超过 $V_U = \min_j \max_i a_{ij}$ 。一般称 V_L 与 V_U 为对策的下、上限值,而在对策中 $V_L = V_U$ 时,此对策有一个鞍点,或称对策有解。

例 11.1 设一对策 $G = \{S, D, A\}$,其中 $S = \{s_1, s_2, s_3\}$, $D = \{d_1, d_2, d_3\}$,其支付矩阵为

$$A = \begin{bmatrix} 3 & 1 & 2 \\ 6 & 0 & -3 \\ -5 & -1 & 4 \end{bmatrix}$$

问:双方局中人采用何策略最佳。

解 可用下述表格表示上述寻找最优纯策略的过程:

	d_1	d_2	d_3	$f(s_i) = \min_j a_{ij}$	
s_1	3	1	2	1	$V_L = \max_i \min_j a_{ij} = 1$
s_2	6	0	-3	-3	
s_3	-5	-1	4	-5	$V_U = \min_j \max_i a_{ij} = 1$
$g(d_j) = \max_i a_{ij}$	6	1	4	$V_L = V_U = a_{12} = 1$	

因为 $V_L = V_U = 1$, 显然此对策有解, 局势 (s_1, d_2) 构成一个鞍点, 局中人甲的最优策略为 s_1 , 局中人乙的最优策略为 d_2 , 对策值 $V = 1$ 。

(2) 多鞍点与无鞍点对策

对于一个纯策略 G , 并非一定只存在惟一鞍点。从下面二例可发现, 对于一个对策有可能有多个鞍点(多个解), 或者根本不存在鞍点, 即无纯策略解的情况。

例 11.2 设有一矩阵对策为

$$A = \begin{bmatrix} 6 & 5 & 6 & 5 \\ 1 & 4 & 2 & -1 \\ 8 & 5 & 7 & 5 \\ 0 & 2 & 6 & 2 \end{bmatrix}$$

求它的解。

解

	d_1	d_2	d_3	d_4	$f(s_i)$
s_1	6	5	6	5	5
s_2	1	4	2	-1	-1
s_3	8	5	7	5	5
s_4	0	2	6	2	0
$g(d_j)$	8	5	7	5	5

$$V = \min_j \max_i a_{ij} = \max_i \min_j a_{ij} = 5$$

对策有解, 解值 $V = 5$, 但局势 (s_1, d_2) 、 (s_1, d_4) 、 (s_3, d_2) 、 (s_3, d_4) 均构成鞍点。此对策有多个解。

例 11.3 矩阵对策支付矩阵

$$A = \begin{bmatrix} 1 & 0 \\ -4 & 3 \end{bmatrix}$$

求它的解。

解

$$V_L = \max_i \min_j a_{ij} = \max\{0, -4\} = 0$$

$$V_U = \min_j \max_i a_{ij} = \min\{1, 3\} = 1$$

$V_L < V_U$, 则对策无鞍点, 即无解。

(3) 优超原理

对于一个对策 $G = \{S, D, A\}$, 对 $s_k \in S$ 至少存在一个策略 $s_i \in S$, 满足

$$a_{ij} \geq a_{kj} \quad (j=1, \dots, n)$$

称 s_k 为局中人甲的一个劣策略。因为在任何情况下, 策略 s_i 都优于策略 s_k , 则在对策分析中

可不考虑 s_k , 且可从策略集 S 中取消它。同时, 可在支付矩阵 A 中删去第 k 行支付值。同样, 对局中人乙也如此。若策略 d_j 与 d_k 满足

$$a_{ik} \geq a_{ij} \quad (i=1, \dots, m)$$

认为策略 d_j 优于 d_k , 可从 D 中删去 d_k , 并从 A 矩阵中删去第 k 列支付值。用此方法来降低支付矩阵的维数, 从而简化问题, 称此为优超原理。

例 11.4 利用优超原理求解下列对策。

$$\begin{array}{c} d_1 \quad d_2 \quad d_3 \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \left[\begin{array}{ccc} 0 & 2 & 1 \\ 2 & 2 & 3 \\ 1 & 3 & 3 \end{array} \right] \end{array}$$

解 首先利用优超原理将上述矩阵对策降维

$$\left[\begin{array}{ccc} 0 & 2 & 1 \\ 2 & 2 & 3 \\ 1 & 3 & 3 \end{array} \right] \xrightarrow{s_1 \succ s_3} \begin{array}{c} d_1 \quad d_2 \quad d_3 \\ \begin{array}{c} s_2 \\ s_3 \end{array} \left[\begin{array}{ccc} 2 & 2 & 3 \\ 1 & 3 & 3 \end{array} \right] \end{array} \xrightarrow{d_2 \succ d_3} \begin{array}{c} d_1 \quad d_2 \\ \begin{array}{c} s_2 \\ s_3 \end{array} \left[\begin{array}{cc} 2 & 2 \\ 1 & 3 \end{array} \right] \end{array} = A$$

其中的符号 \succ 表示优于。对于 A , $\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = 2$, 即对策解或鞍点为 (s_2, d_1) , 对策的值为 2。

11.3 二人有限零和对策的混合策略对策模型

11.3.1 混合策略对策的基本概念

在一局纯策略矩阵对策中没有鞍点, 或者对策要进行多次时, 任一方坚持采用一种固定的策略是不明智的, 如同在上节所述的中国儿童的对策游戏“猜手”。如果进行多次, 哪一个孩子也不会只出一种手势, 而是有时出“剪子”, 有时出“布”, 有时出“石头”。所以对无鞍点对策 ($V_L \neq V_U$) 的求解方法是采用混合策略。混合策略就是局中人考虑以某种概率分布来选择他的各个策略。

一个纯策略对策 $G = (S, D, A)$, $S = \{s_1, s_2, \dots, s_m\}$, $D = \{d_1, d_2, \dots, d_n\}$ 分别为局中人甲、乙的一个策略集合, $A = [a_{ij}]_{m \times n}$ 为得失矩阵。

1. 混合策略

m 维概率向量 $x = [x_1 \quad x_2 \quad \dots \quad x_m]^T$, $\sum_{i=1}^m x_i = 1, x_i \geq 0 (i=1, 2, \dots, m)$ 称为局中人甲的一个混合策略, 即局中人甲选择策略 s_i 的概率为 x_i 。 n 维概率向量 $y = [y_1 \quad y_2 \quad \dots \quad y_n]^T$, $\sum_{j=1}^n y_j = 1, y_j \geq 0 (j=1, \dots, n)$ 称为乙的一个混合策略。局中人乙选择策略 d_j 的概率为 $y_j (j=1, 2, \dots, n)$ 。

2. 混合策略集合

人们称集合

$S^* = \left\{ x = [x_1 \ x_2 \ \cdots \ x_m]^T \mid \sum_{i=1}^m x_i = 1, x_i \geq 0 \right\}$ 为甲的混合策略集合;

$D^* = \left\{ y = [y_1 \ y_2 \ \cdots \ y_n]^T \mid \sum_{j=1}^n y_j = 1, y_j \geq 0 \right\}$ 为乙的混合策略集合。

3. 混合局势

当局中人甲选择混合策略 x 、乙选择混合策略 y 时,称 (x, y) 为一个混合局势。

4. 收益期望值

对于一个混合局势 (x, y) , 用

$$E(x, y) = \sum_{i=1}^m x_i \left(\sum_{j=1}^n a_{ij} y_j \right) = x^T A y$$

表示局中人甲在混合局势 (x, y) 时的收益期望值。

5. 混合策略对策模型

对于一个纯策略对策 $G = (S, D, A)$, 用 $G^* = (S^*, D^*, E)$ 表示一个与之相应的混合策略矩阵对策, 也称 G 的一个混合扩充。

11.3.2 混合策略对策的解

1. 混合策略分析

混合策略对策 $G^* = (S^*, D^*, E)$ 仍然是一个竞争性决策问题。局中人在选择混合策略时, 也首先得对其各混合策略进行评价, 在评价时应考虑对方可能采用的混合策略。

局中人甲选择混合策略 x 时, 认为乙必采用使甲期望收益最小的混合策略 y_x , 所以局中人甲对其混合策略的评价函数为

$$\begin{aligned} f(x) &= \min_{y \in D^*} (x, y) = \min_{y \in D^*} x^T A y \\ f(x) &= \min_y E(x, y) = E(x, y_x) \end{aligned}$$

局中人甲的策略决策模型为

$$\max_{x \in S^*} f(x) = \max_{x \in S^*} \min_{y \in D^*} E(x, y) = E(x^*, y_x^*) \rightarrow x^*$$

同样, 局中人乙若采用混合策略 y , 认为甲必采用使甲期望收益最大的混合策略, 所以局中人乙对其混合策略的评价函数为

$$\begin{aligned} g(y) &= \max_{x \in S^*} E(x, y) = \max_{x \in S^*} x^T A y \\ g(y) &= \max_x E(x, y) = E(x_y, y) \end{aligned}$$

局中人乙的策略决策模型为

$$\min_{y \in D^*} g(y) = \min_{y \in D^*} \max_{x \in S^*} E(x, y) = E(x_y^*, y^*) \rightarrow y^*$$

容易证明下列定理。

定理 11.2 设 $G^* = (S^*, D^*, E)$ 是混合策略对策, 那么必有

$$E(x_y^*, y^*) = \min_{y \in D^*} \max_{x \in S^*} E(x, y) \geq \max_{x \in S^*} \min_{y \in D^*} E(x, y) = E(x^*, y_x^*)$$

证明 因为

$$\begin{aligned} f(x) &\leq E(x, y) \leq g(y) \\ \min_y E(x, y) &\leq E(x, y) \leq \max_x E(x, y) \end{aligned}$$

所以

$$E(x^*, y_x^*) = \max_x f(x) \leq \min_y g(y) = E(x_y^*, y^*)$$

2. 混合策略对策的解

如果混合策略对策 $G^* = (S^*, D^*, E)$ 满足

$$E(x^*, y_x^*) = \max_{x \in S^*} \min_{y \in D^*} E(x, y) = \min_{y \in D^*} \max_{x \in S^*} E(x, y) = E(x_y^*, y^*)$$

根据定理 11.2, 必有

$$E(x^*, y_x^*) = E(x^*, y^*) = E(x_y^*, y^*)$$

则称混合局势 (x^*, y^*) 为 G^* 的一个鞍点。 x^* 为甲的最优混合策略, y^* 为乙的最优混合策略, $V = E(x^*, y^*)$ 为混合策略对策 G^* 的值, (x^*, y^*) 为 G^* 的一个解。

3. 混合策略对策的线性规划解法

对于混合策略对策 $G^* = (S^*, D^*, E)$, 局中人甲是通过模型

$$E(x^*, y_x^*) = \max_{x \in S^*} \min_{y \in D^*} E(x, y)$$

确定 $x^* = [x_1^* \ x_2^* \ \cdots \ x_m^*]^T$ 。而其中 $\min_{y \in D^*} E(x, y) = \min_{y \in D^*} \sum_{j=1}^n (\sum_{i=1}^m a_{ij} x_i) y_j$, 令

$$\sum_{i=1}^m a_{ij} x_i = b_j$$

则上式化为

$$\min E(x, y) = \min_{y \in D^*} \sum_{j=1}^n b_j y_j = u$$

显然 $b_j \geq u$ ($j=1, 2, \dots, n$) 且 $u = \min_j b_j$ 。

上述求解 x^* 问题相当于下列线性规划问题:

$$\begin{aligned} & \max u \\ & \text{s. t. } \begin{cases} \sum_{i=1}^m a_{ij} x_i \geq u & (j=1, 2, \dots, n; i=1, \dots, m) \\ \sum_{i=1}^m x_i = 1 \\ x_i \geq 0, \end{cases} \end{aligned} \quad (11.1)$$

$\bar{x} = [1 \ 0 \ \cdots \ 0]^T$, $\bar{u} = \min_j a_{ij}$ 显然是此线性规划问题的一个可行解。同样对局中人乙

确定 y^* 的模型

$$E(x_y^*, y^*) = \min_{y \in D^*} \max_{x \in S^*} E(x, y) \rightarrow y^* = [y_1^* \ y_2^* \ \cdots \ y_n^*]^T$$

可转换为求解下列线性规划问题:

$$\begin{aligned} & \min v \\ & \text{s. t. } \begin{cases} \sum_{j=1}^n a_{ij} y_j \leq v & (i=1, 2, \dots, m; j=1, 2, \dots, n) \\ \sum_{j=1}^n y_j = 1 \\ y_j \geq 0 \end{cases} \end{aligned} \quad (11.2)$$

同样 $\bar{y} = [1 \ 0 \ \cdots \ 0]^T$, $\bar{v} = \max_i a_{ij}$ 为此问题的可行解。

不失一般性, 假设得失矩阵 $A = [a_{ij}]_{m \times n}$ 中 $a_{ij} > 0$, 否则可取一充分大的正数 $M > 0$, 使得

$$a'_{ij} = a_{ij} + M > 0 \quad (i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$$

并构成新的得失矩阵

$$A' = [a'_{ij}]_{m \times n}$$

和纯策略矩阵对策

$$G_1 = (S, D, A')$$

对于 $G = (S, D, A)$ 和 $G_1 = (S, D, A')$ 及对应的混合策略对策 $G^* = (S^*, D^*, E)$ 和 $G_1^* = (S^*, D^*, E')$, 考虑它们的期望收益函数 $E(x, y)$ 和 $E_1(x, y)$, 有

$$\begin{aligned} E_1(x, y) &= \sum_{i=1}^m \sum_{j=1}^n a'_{ij} x_i y_j = \sum_{i=1}^m \sum_{j=1}^n (a_{ij} + M) x_i y_j \\ &= \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j + M \cdot \sum_{i=1}^m \sum_{j=1}^n x_i y_j \\ &= E(x, y) + M \end{aligned}$$

由此可得以下定理。

定理 11.3 若 G^* 有解 (x^*, y^*) , 则 G_1^* 也有解 (x^*, y^*) , 且对策值相差 M , 即 $V_1 = V + M$ 。

在 $a_{ij} > 0$ 的假定条件下, 对于线性规划(11.1)有一可行解

$$x = [1 \ 0 \ \cdots \ 0]^T, \bar{u} = \min_j a_{ij} > 0$$

所以, 必有 $u > 0$ 。同样, 对线性规划式(11.2)的可行解中, 必满足

$$v \geq \min_i a_{ij} > 0$$

对这样两个线性规划问题, 令

$$x'_i = \frac{x_i}{u} \quad (i = 1, 2, \cdots, m); y'_j = \frac{y_j}{v} \quad (j = 1, 2, \cdots, n)$$

有

$$\sum_{i=1}^m x'_i = \frac{1}{u}, \sum_{j=1}^n y'_j = \frac{1}{v}$$

$$x' = \frac{x}{u}, y' = \frac{y}{v}$$

可将上述两规划转化为下述形式:

$$\begin{aligned} \min \quad & w = x'_1 + x'_2 + \cdots + x'_m \\ \text{s. t.} \quad & \begin{cases} \sum_{i=1}^m a_{ij} x'_i \geq 1 & (j = 1, 2, \cdots, n) \\ x'_i \geq 0 & (i = 1, 2, \cdots, m) \end{cases} \\ \max \quad & z = y'_1 + y'_2 + \cdots + y'_n \\ \text{s. t.} \quad & \begin{cases} \sum_{j=1}^n a_{ij} y'_j \leq 1 & (i = 1, 2, \cdots, m) \\ y'_j \geq 0 & (j = 1, 2, \cdots, n) \end{cases} \\ x^* &= \frac{1}{w^*} x', y^* = \frac{1}{z^*} y' \end{aligned}$$

显然有

$$w^* = \frac{1}{u^*} = \frac{1}{\max_x \min_y E(x, y)} = \frac{1}{E(x^*, y^*)}$$

$$z^* = \frac{1}{v^*} = \frac{1}{\min_y \max_x E(x, y)} = \frac{1}{E(x^*, y^*)}$$

而此两线性规划问题是互为对偶的, $w^* = z^*$, 则有

$$\max_x \min_y E(x, y) = \min_y \max_x E(x, y) = E(x^*, y^*)$$

(x^*, y^*) 为 G^* 的一个解。

4. 混合策略矩阵对策解的存在性

使用线性规划方法求解 $G^* = (S^*, D^*, E)$ 时, 由上述分析得出下面的结论。

定理 11.4 对于 $G^* = (S^*, D^*, E)$, 必存在最优解 (x^*, y^*) , 满足

$$E(x^*, y^*) = \max_x \min_y E(x, y) = \min_y \max_x E(x, y)$$

例 11.5 求解下列混合策略矩阵对策 $G^* = (S^*, D^*, E)$, 其中

$$S = \{s_1, s_2, s_3\}; D = \{d_1, d_2\}$$

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 0 & 1 \end{bmatrix}。$$

解 令

$$A' = [a_{ij} + 2] = \begin{bmatrix} 3 & 1 \\ 1 & 4 \\ 2 & 3 \end{bmatrix}$$

确定甲的最优混合策略, 求解下述线性规划

$$\begin{aligned} \min \quad & w = x_1 + x_2 + x_3 \\ \text{s. t.} \quad & \begin{cases} 3x_1 + x_2 + 2x_3 \geq 1 \\ x_1 + 4x_2 + 3x_3 \geq 1 \\ x_j \geq 0, j = 1, 2, 3 \end{cases} \end{aligned}$$

使用单纯形法, 得 $x' = \left[\frac{1}{7} \quad 0 \quad \frac{2}{7} \right]^T$, $w^* = \frac{3}{7}$, 同时也可得其对偶问题解: $y' = \left[\frac{2}{7} \quad \frac{1}{7} \right]^T$, $z^* = \frac{3}{7}$ 。

由于 $x' = \frac{1}{w} x^*$, $y' = \frac{1}{z} y^*$, 可得最优策略

$$x^* = \frac{1}{w} x' = \frac{7}{3} \left[\frac{1}{7} \quad 0 \quad \frac{2}{7} \right]^T = \left[\frac{1}{3} \quad 0 \quad \frac{2}{3} \right]^T$$

$$y^* = \frac{1}{z} y' = \frac{7}{3} \left[\frac{2}{7} \quad \frac{1}{7} \right]^T = \left[\frac{2}{3} \quad \frac{1}{3} \right]^T$$

例 11.6 设

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 2 & 3 \\ 1 & 3 & 3 \end{bmatrix}$$

试求解 $G^* = (S^*, D^*, A)$ 。

解 先用优超原理降低 A 的维数:

$$A = \begin{matrix} & d_1 & d_2 & d_3 \\ \begin{matrix} s_1 \\ s_2 \\ s_3 \end{matrix} & \begin{bmatrix} 0 & 2 & 1 \\ 2 & 2 & 3 \\ 1 & 3 & 3 \end{bmatrix} \end{matrix} \longrightarrow \begin{bmatrix} 0 & 2 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} \longrightarrow \begin{matrix} & d_1 & d_2 \\ \begin{matrix} s_2 \\ s_3 \end{matrix} & \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix} \end{matrix} = A'$$

化为 $G' = (S', D', A')$, 其中 $S' = \{s_2, s_3\}$, $D' = \{d_1, d_2\}$ 。现求解 $G^* = (S'^*, D'^*, E')$ 。

求乙方最优策略 y^* , 即求解线性规划问题:

$$\begin{aligned} \max \quad & z = y_1 + y_2 \\ \text{s. t.} \quad & \begin{cases} y_1 + 2y_2 \leq 1 \\ y_1 + y_2 \leq 1 \\ y_j \geq 0 \quad (j=1, 2) \end{cases} \end{aligned}$$

最优解 $y' = [y'_1 \quad y'_2]^T = \left[\frac{1}{5} \quad \frac{2}{5}\right]^T$, $z^* = \frac{3}{5}$, 其对偶解为

$$x' = [x'_1 \quad x'_2]^T = \left[\frac{2}{5} \quad \frac{1}{5}\right]^T, w^* = \frac{3}{5}$$

最优策略 $x^* = \frac{1}{w^*} [x'_1 \quad x'_2 \quad x'_3]^T = \frac{5}{3} \left[\frac{2}{5} \quad 0 \quad \frac{1}{5}\right]^T = \left[\frac{2}{3} \quad 0 \quad \frac{1}{3}\right]^T$

$$y^* = \frac{1}{z^*} [y'_1 \quad y'_2 \quad y'_3]^T = \frac{5}{3} \left[\frac{1}{5} \quad \frac{2}{5} \quad 0\right]^T = \left[\frac{1}{3} \quad \frac{2}{3} \quad 0\right]^T$$

G^* 解值 $u = \frac{5}{3}$ 。

5. 当矩阵为 2×2 阶时解的公式

当局中人甲的支付矩阵 A 为 2×2 的简单情形时, 若鞍点不存在, 则容易得出下面的解的公式。

$$\begin{aligned} \text{记 } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ 则 } (x^*, y^*, v^*) = & \left(\left[\frac{d-c}{a-b-c+d} \quad \frac{a-b}{a-b-c+d} \right]^T, \right. \\ & \left. \left[\frac{d-b}{a-b-c+d} \quad \frac{a-c}{a-b-c+d} \right]^T, \frac{ad-bc}{a-b-c+d} \right) \end{aligned} \quad (11.3)$$

例 11.7 求解矩阵对策 $G = (S, D, A)$, 其中 $A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$ 。

解 由 (11.3) 式, $x^* = \left[\frac{2-4}{1-3-4+2} \quad \frac{1-3}{1-3-4+2} \right]^T = \left[\frac{1}{2} \quad \frac{1}{2} \right]^T$, $y^* = \left[\frac{2-3}{1-3-4+2} \quad \frac{1-4}{1-3-4+2} \right]^T = \left[\frac{1}{4} \quad \frac{3}{4} \right]^T$, $v^* = \frac{1 \times 2 - 3 \times 4}{1-3-4+2} = \frac{10}{4} = \frac{5}{2}$ 。

习 题

11.1 甲、乙两个企业生产同一种电子产品, 两个企业都想通过改革管理获取更多的市场销售份额。甲企业的策略措施有: ①降低产品价格; ②提高产品质量, 延长保修年限; ③推出新产品。乙企业考虑的措施有: ①增加广告费用; ②增设维修网点, 扩大维修服务; ③改进产品性能。假定市场容量一定, 由于各自采取的策

略措施不同,通过预测,今后两个企业的市场占有率变动情况如表 11.1 所示(正值为甲企业增加的市场占有份额,负值为甲企业减少的市场占有份额)。试通过对策分析,确定两个企业各自的最优策略。

表 11.1

		乙企业策略		
		1	2	3
甲企业策略	1	10	-1	3
	2	12	10	-5
	3	6	8	5

11.2 有 A、B 两家生产小型电子计算器的工厂,其中 A 厂研制出一种新型袖珍计算器。为推出这种新产品加强与 B 厂竞争,考虑了三个竞争对策:①将新产品全面投入生产;②继续生产现有产品,新产品小批量试产试销;③维持原状,新产品只生产样品征求意见。B 厂了解到 A 厂有新产品情况下也考虑了三个策略:①加速研制新计算器;②对现有计算器革新;③改进产品外观和包装。由于受市场预测能力限制,表 11.2 只表明双方对策结果的大致的定性分析资料(对 A 厂而言):

若用打分办法,一般记 0 分,较好记 1 分,好记 2 分,很好记 3 分,较差记 -1 分,差记 -2 分,很差记 -3 分,试通过对策分析,确定 A、B 两厂各应采取哪一种策略。

表 11.2

		B 厂策略		
		1	2	3
A 厂策略	1	较好	好	很好
	2	一般	较差	较好
	3	很差	差	一般

11.3 已知矩阵对策

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 8 \\ 0 & 6 & 0 \end{bmatrix}$$

的解为 $x^* = [6/13 \ 3/13 \ 4/13]^T$, $y^* = [6/13 \ 4/13 \ 3/13]^T$, 对策值为 $24/13$ 。求下列矩阵对策的解,其支付矩阵 A 分别为

$$(a) \begin{bmatrix} 6 & 2 & 2 \\ 2 & 2 & 10 \\ 2 & 8 & 2 \end{bmatrix} \quad (b) \begin{bmatrix} -2 & -2 & 2 \\ 6 & -2 & -2 \\ -2 & 4 & -2 \end{bmatrix}$$

11.4 用线性规划方法求解下面的矩阵对策,其中 A 为

$$(a) \begin{bmatrix} 8 & 2 & 4 \\ 2 & 6 & 6 \\ 6 & 4 & 4 \end{bmatrix} \quad (b) \begin{bmatrix} 2 & 0 & 2 \\ 0 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix}$$

第 12 章 二人有限非零和对策

在二人有限零和对策中,由于两个局中人甲和乙的得与失之和恰为零,故在对问题的一般描述时只需考虑甲的支付矩阵 A 。因此,二人有限零和对策又称为矩阵对策。而在二人的得与失之和一般不为零的非零和情形,对问题的一般描述就必须同时考虑甲的支付矩阵 A 和乙的支付矩阵 B ,因此,二人有限非零和对策又称为双矩阵对策。

双矩阵对策记为 $G = (S, D, (A, B))$, 其中 $S = \{s_1, \dots, s_m\}$ 为局中人甲的策略集, $D = \{d_1, \dots, d_n\}$ 为局中人乙的策略集, $A = [a_{ij}]_{m \times n}$ 和 $B = [b_{ij}]_{m \times n}$ 分别为甲和乙的支付矩阵, a_{ij} 和 b_{ij} 分别表示甲和乙的相应于 s_i 和 d_j 的赢得。

显然,当 $A + B = 0$ 时,双矩阵对策即化为矩阵对策。

在矩阵对策中,由于甲的得就是乙的失,二人之间没有共同利益,故二人是处于完全竞争的非合作状态。但在双矩阵对策中,由于甲的得并不一定等于乙的失,二人可以同时得,故二人之间有可能合作,从而得到更多的利益。

双矩阵对策的求解一般要比矩阵对策复杂得多。本章主要介绍比较简单的 2×2 阶矩阵的情形。

12.1 非合作的二人有限非零和对策

12.1.1 解的概念和存在性定理

与矩阵对策的情形类似,双矩阵对策一般也要在混合策略意义下求解。仍记 $S^* = \{x = [x_1 \ \dots \ x_m]^T \mid x_i \geq 0, \sum_{i=1}^m x_i = 1\}$ 和 $D^* = \{y = [y_1 \ \dots \ y_n]^T \mid y_j \geq 0, \sum_{j=1}^n y_j = 1\}$ 分别为局中人甲和乙的混合策略集。

定义 12.1 设 $x^* \in S^*, y^* \in D^*$, 若对任何 $x \in S^*$ 和任何 $y \in D^*$, 有

$$x^T A y^* \leq x^{*T} A y^*$$

$$x^{*T} B y \leq x^{*T} B y^*$$

则称 (x^*, y^*) 为双矩阵对策 G 的平衡局势。

上述定义给出了双矩阵对策的一种最常用的解的概念,平衡局势 (x^*, y^*) 就是 G 的解,它相应的二局中人的期望收益 $(x^{*T} A y^*, x^{*T} B y^*)$ 就是 G 的值,记为 (U^*, V^*) 。显然,当 $A + B = 0$ 时,平衡局势的定义就化为矩阵对策在混合策略意义下的解的定义了。纳什(Nash)证明了平衡局势的存在性定理。

定理 12.1 任何双矩阵对策至少存在一个平衡局势。

还可证明下面的结论。

定理 12.2 (x^*, y^*) 为双矩阵对策 G 的一个平衡局势的充要条件是存在数 p^* 和 q^* 使 $[x^* \ y^* \ p^* \ q^*]^T$ 是下述规划问题的一个解:

$$\max(x^T A y + x^T B y - p - q)$$

$$\text{s. t. } \begin{cases} Ay \leq pE_n \\ x^T B \leq qE_m \\ E_n^T x = E_m^T y = 1 \\ x \geq 0, y \geq 0 \end{cases}$$

其中 E_n 和 E_m 是分量均为 1 的 n 维和 m 维向量。

定理 12.2 说明, 求解双矩阵对策的问题可以转化为求解一个数学规划问题。但由于这是一个非线性规划, 一般来说求解比较复杂。不过当矩阵 A 和 B 均为 2×2 阶时, 可以利用较简单的方法求解。

12.1.2 2×2 双矩阵对策的解法

当矩阵 A 和 B 均为 2×2 阶时, 相应的双矩阵对策可表示为

$$\begin{array}{cc} & \begin{matrix} \text{乙} \\ y_1 & 1-y_1 \end{matrix} \\ \begin{matrix} \text{甲} \\ x_1 & 1-x_1 \end{matrix} & \begin{bmatrix} (a_{11}, b_{11}) & (a_{12}, b_{12}) \\ (a_{21}, b_{21}) & (a_{22}, b_{22}) \end{bmatrix} \end{array} \quad (0 \leq x_1 \leq 1, 0 \leq y_1 \leq 1)$$

若 (x^*, y^*) 是均衡局势, 由定义 12.1, x^* 和 y^* 应分别是 $x^T Ay^*$ 在 S^* 上和 $x^{*T} By$ 在 D^* 上的极大点。而

$$\begin{aligned} x^T Ay^* &= [x_1 \quad 1-x_1] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_1^* \\ 1-y_1^* \end{bmatrix} \\ &= [(a_{11} - a_{12} - a_{21} + a_{22})y_1^* - (a_{22} - a_{12})]x_1 + a_{22} + (a_{21} - a_{22})y_1^* \end{aligned}$$

记

$$A_1 = a_{11} - a_{12} - a_{21} + a_{22}$$

$$A_2 = a_{22} - a_{12}$$

则使 $x^T Ay^*$ 达极大的 x_1^* 应满足

$$x_1^* = \begin{cases} 0, & \text{当 } A_1 y_1^* - A_2 < 0 \\ [0, 1] \text{ 中任意值,} & \text{当 } A_1 y_1^* - A_2 = 0 \\ 1, & \text{当 } A_1 y_1^* - A_2 > 0 \end{cases} \quad (12.1)$$

$$\begin{aligned} x^{*T} By &= [x_1^* \quad 1-x_1^*] \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ 1-y_1 \end{bmatrix} \\ &= [(b_{11} - b_{12} - b_{21} + b_{22})x_1^* - (b_{22} - b_{21})]y_1 + b_{22} + (b_{12} - b_{22})x_1^* \end{aligned}$$

记

$$B_1 = b_{11} - b_{12} - b_{21} + b_{22}$$

$$B_2 = b_{22} - b_{21}$$

则使 $x^{*T} By$ 达极大的 y_1^* 应满足

$$y_1^* = \begin{cases} 0, & \text{当 } B_1 x_1^* - B_2 < 0 \\ [0, 1] \text{ 中任意值,} & \text{当 } B_1 x_1^* - B_2 = 0 \\ 1, & \text{当 } B_1 x_1^* - B_2 > 0 \end{cases} \quad (12.2)$$

根据式(12.1)和式(12.2),可在以 x_1 和 y_1 为横、纵轴的坐标系中确定出对于局中人甲来说可能成为平衡局势的点(不妨称为甲的解) (x_1^*, y_1^*) 的轨迹和对于局中人乙来说可能成为平衡局势的点(不妨称为乙的解) (x_1, y_1^*) 的轨迹。二轨迹的公共点即 (x_1^*, y_1^*) , 由此便可得到平衡局势 (x^*, y^*) 。将这一分析的结果各分为 9 种情形(即 A_1 和 A_2 、 B_1 和 B_2 取各种符号时的 9 种条件)列于表 12.1 和表 12.2。

表 12.1

条件序号	条件	解	图示
1	$A_1 = 0$ $A_2 = 0$	$0 \leq x_1 \leq 1$ $(0 \leq y_1 \leq 1)$	
2	$A_1 = 0$ $A_2 > 0$	$x_1 = 0$ $(0 \leq y_1 \leq 1)$	
3	$A_1 = 0$ $A_2 < 0$	$x_1 = 1$ $(0 \leq y_1 \leq 1)$	
4	$A_1 > 0$ $A_2 = 0$	$x_1 = 1 (0 \leq y_1 \leq 1)$ $y_1 = 0 (0 \leq x_1 \leq 1)$	
5	$A_1 < 0$ $A_2 = 0$	$x_1 = 0 (0 \leq y_1 \leq 1)$ $y_1 = 0 (0 \leq x_1 \leq 1)$	
6	$A_1 > 0$ $A_2 > 0$	$x_1 = 0 (0 \leq y_1 \leq \frac{A_2}{A_1})$ $y_1 = \frac{A_2}{A_1} (0 \leq x_1 \leq 1)$ $x_1 = 1 (\frac{A_2}{A_1} \leq y_1 \leq 1)$	
7	$A_1 < 0$ $A_2 < 0$	$x_1 = 0 (\frac{A_2}{A_1} \leq y_1 \leq 1)$ $y_1 = \frac{A_2}{A_1} (0 \leq x_1 \leq 1)$ $x_1 = 1 (0 \leq y_1 \leq \frac{A_2}{A_1})$	
8	$A_1 > 0$ $A_2 < 0$	$x_1 = 1$ $(0 \leq y_1 \leq 1)$	
9	$A_1 < 0$ $A_2 > 0$	$x_1 = 0$ $(0 \leq y_1 \leq 1)$	

表 12.2

条件序号	条件	解	图示
1	$B_1 = 0$ $B_2 = 0$	$0 \leq y_1 \leq 1$ $(0 \leq x_1 \leq 1)$	
2	$B_1 = 0$ $B_2 > 0$	$y_1 = 0$ $(0 \leq x_1 \leq 1)$	
3	$B_1 = 0$ $B_2 < 0$	$y_1 = 1$ $(0 \leq x_1 \leq 1)$	
4	$B_1 > 0$ $B_2 = 0$	$y_1 = 1 (0 \leq x_1 \leq 1)$ $x_1 = 0 (0 \leq y_1 \leq 1)$	
5	$B_1 < 0$ $B_2 = 0$	$y_1 = 0 (0 \leq x_1 \leq 1)$ $x_1 = 0 (0 \leq y_1 \leq 1)$	
6	$B_1 > 0$ $B_2 > 0$	$y_1 = 0 (0 \leq x_1 \leq \frac{B_2}{B_1})$ $x_1 = \frac{B_2}{B_1} (0 \leq y_1 \leq 1)$ $y_1 = 1 (\frac{B_2}{B_1} \leq x_1 \leq 1)$	
7	$B_1 < 0$ $B_2 < 0$	$y_1 = 0 (\frac{B_2}{B_1} \leq x_1 \leq 1)$ $x_1 = \frac{B_2}{B_1} (0 \leq y_1 \leq 1)$ $y_1 = 1 (0 \leq x_1 \leq \frac{B_2}{B_1})$	
8	$B_1 > 0$ $B_2 < 0$	$y_1 = 1$ $(0 \leq x_1 \leq 1)$	
9	$B_1 < 0$ $B_2 > 0$	$y_1 = 0$ $(0 \leq x_1 \leq 1)$	

总结 2×2 阶双矩阵对策的求解步骤如下:

①由矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

计算

$$A_1 = a_{11} - a_{12} - a_{21} + a_{22}, A_2 = a_{22} - a_{12}$$

$$B_1 = b_{11} - b_{12} - b_{21} + b_{22}, B_2 = b_{22} - b_{21}$$

②根据 A_i 和 B_i ($i=1,2$) 的符号,由表 12.1 和表 12.2 得到局中人甲和乙的解,其公共点即对策的解。

例 12.1(夫妇之争) 夫妇俩商量晚上到哪儿去消遣。丈夫喜欢去看足球比赛,而妻子喜欢去看芭蕾舞表演,夫妇二人都希望二人同往。

解 现将该问题归为一个双矩阵对策。记丈夫为局中人甲,其策略为 s_1 (看芭蕾)、 s_2 (看足球),相应混合策略 $(x_1, 1-x_1)$;妻子为局中人乙,其策略为 d_1 (看芭蕾)、 d_2 (看足球),相应混合策略 $(y_1, 1-y_1)$ 。设其得失矩阵为

$$\begin{array}{c} \text{乙} \\ \text{甲} \begin{bmatrix} (1,4) & (0,0) \\ (0,0) & (4,1) \end{bmatrix} \end{array}$$

矩阵中位于第 i 行 j 列的括号即 (a_{ij}, b_{ij}) , 则

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A_1 = 1 - 0 - 0 + 4 = 5, A_2 = 4 - 0 = 4, \frac{A_2}{A_1} = \frac{4}{5}$$

$$B_1 = 4 - 0 - 0 + 1 = 5, B_2 = 1 - 0 = 1, \frac{B_2}{B_1} = \frac{1}{5}$$

丈夫的解(图 12.1)为

$$x_1 = 0, 0 \leq y_1 \leq \frac{4}{5}$$

$$x_1 = 1, \frac{4}{5} \leq y_1 \leq 1$$

$$y_1 = \frac{4}{5}, 0 \leq x_1 \leq 1$$

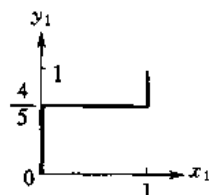


图 12.1

妻子的解(图 12.2)为

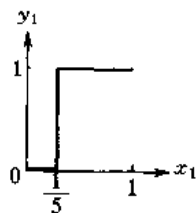


图 12.2

$$y_1 = 0, 0 \leq x_1 \leq \frac{1}{5}$$

$$0 \leq y_1 \leq 1, x_1 = \frac{1}{5}$$

$$y_1 = 1, \frac{1}{5} \leq x_1 \leq 1$$

由解的图示易见,公共点有三个: $A(0,0)$, $B(\frac{1}{5}, \frac{4}{5})$, $C(1,1)$, 如图 12.3 所示。

A 点相应于平衡局势 $(x^*, y^*) = ([0 \ 1]^T, [0 \ 1]^T)$, 即二人同去看足球, 相应的得失值 $(U^*, V^*) = (x^{*T} A y^*, x^{*T} B y^*) = (4, 1)$, 丈夫得到最大的满足。

B 点相应于平衡局势 $(x^*, y^*) = ([\frac{1}{5} \ \frac{4}{5}]^T, [\frac{4}{5} \ \frac{1}{5}]^T)$, 相应的得失值 $(U^*, V^*) =$

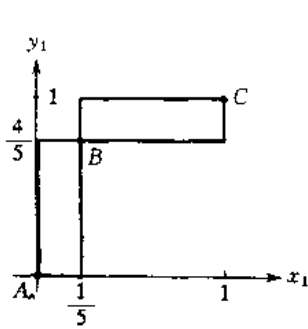


图 12.3

$\left(\frac{4}{5}, \frac{4}{5}\right)$, 两人在此混合策略下的满意度得到了均衡, 但却都降低了。

C 点相应于平衡局势 $(x^*, y^*) = ((1,0), (1,0))$, 即二人同去看芭蕾, 相应得失值 $(U^*, V^*) = (1,4)$, 妻子得到最大的满足。

由此例可以看到, 双矩阵对策的不同的解可以对应于不同的值, 这一点是与矩阵对策不同的。

例 12.2(囚犯两难推理) 两名囚犯因涉嫌抢劫被捕。警方因证据不足先将二人分关二室, 并宣布: 若二人均不坦白, 则只能因藏有枪支而被判刑 1 年; 若有一人坦白而另一人不坦白, 则坦白者无罪释放, 不坦白者被判刑 10 年; 若二人都坦白了, 则同判刑 9 年。此二人确系抢劫犯。请分析他们的抉择。

解 将此问题看做如下的双矩阵对策问题:

		囚犯乙	
		坦白	不坦白
囚犯甲	坦白	$(-9, -9)$	$(0, -10)$
	不坦白	$(-10, 0)$	$(-1, -1)$

双方的矩阵

$$A = \begin{bmatrix} -9 & 0 \\ -10 & -1 \end{bmatrix}, B = \begin{bmatrix} -9 & -10 \\ 0 & -1 \end{bmatrix}$$

$$A_1 = -9 - 0 - (-10) + (-1) = 0, A_2 = -1 - 0 = -1$$

$$B_1 = -9 - (-10) - 0 + (-1) = 0, B_2 = -1 - 0 = -1$$

囚犯甲的解(如图 12.4)

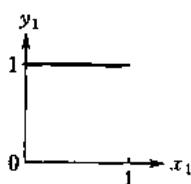


图 12.5

$$x_1 = 1, 0 \leq y_1 \leq 1$$

囚犯乙的解(如图 12.5)

$$y_1 = 1, 0 \leq x_1 \leq 1$$

公共点为 $(x_1, y_1) = (1, 1)$, 相应的均衡局势为 $(x^*, y^*) = ((1,0), (1,0))$, 即二人都坦白, 得失值 $(U^*, V^*) = (-9, -9)$, 即二人均被判刑 9 年。

试想, 如果二囚犯之间允许通话, 那么结果将会怎样呢?

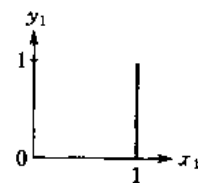


图 12.4

12.2 合作的二人有限非零和对策

前面已经指出, 在非零和对策中, 二局中人有可能进行合作, 共同控制对策的进程, 分享共同利益。以例 12.1 的夫妇之争问题为例, 如果允许合作, 则二人显然不会考虑得失值为 $(0,0)$ 的策略, 而是协商怎样在得失值 $(1,4)$ 和 $(4,1)$ 中进行权衡, 也可以说以多大的概率 α 选取 $(1,4)$ 而以概率 $(1-\alpha)$ 选取 $(4,1)$, 从而得到期望得失值

$$\alpha(1,4) + (1-\alpha)(4,1)$$

这时的对策问题称做采用联合随机策略的合作双矩阵对策。

在非合作的情形,丈夫的期望得失值表达式为

$$U = [x_1 \quad 1-x_1] \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} y_1 \\ 1-y_1 \end{bmatrix} = 5x_1y_1 - 4x_1 - 4y_1 + 4$$

妻子的期望得失值表达式为

$$V = [x_1, 1-x_1] \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ 1-y_1 \end{bmatrix} = 5x_1y_1 - x_1 - y_1 + 1$$

当 x_1 和 y_1 各自在 $[0, 1]$ 上变动时, $U-V$ 平面上的点 (U, V) 构成非合作对策的“可行收益”区域,如图 12.6 中阴影部分所示。

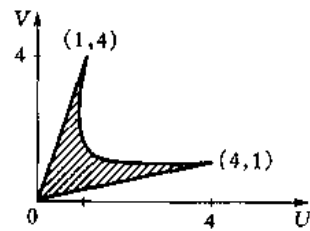


图 12.6

而在合作的情形,使用联合随机策略,以概率 α 选取 $(1, 4)$, 以概率 $(1-\alpha)$ 选取 $(4, 1)$, 所得的得失值 $\alpha(1, 4) + (1-\alpha)(4, 1)$ 表示连接 $(1, 4)$ 和 $(4, 1)$ 的线段,这说明合作时的可行得失域要比非合作时的大。如果以概率 $\alpha_1, \alpha_2, \alpha_3$ ($\alpha_1 + \alpha_2 + \alpha_3 = 1$) 分别选取 $(1, 4)$ 、 $(4, 1)$ 和 $(0, 0)$, 则得失值 $\alpha_1(1, 4) + \alpha_2(4, 1) + \alpha_3(0, 0)$ 将充满以三点为顶点的三角形,如图 12.7 所示。

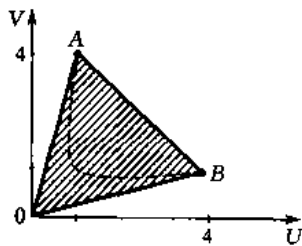


图 12.7

为了得到最大的收益,显然应当在线段 AB 上选取最优的 (U, V) 。因为 AB 上的点 (U, V) 均满足:找不到可行得失域中的点 (U', V') , 使 $U' \geq U$ 且 $V' \geq V$, AB 上的点称为是帕累托 (Pareto) 最优的。

一般地,如果在域中不存在 $(U', V') \neq (U, V)$, 使 $U' \geq U$, $V' \geq V$, 称合作双矩阵对策可行得失域中的点 (U, V) 是 Pareto 最优的。

同时,考虑到合作时的收益不能低于不合作时的最低收益,通常取 (U_0, V_0) 作为基点,其中

$$U_0 = \max_{x \in S^*} \min_{y \in D^*} x^T A y$$

$$V_0 = \max_{y \in D^*} \min_{x \in S^*} x^T B y$$

(U_0, V_0) 称为合作双矩阵对策的安全点。称满足 $U \geq U_0$ 且 $V \geq V_0$ 的 Pareto 最优点 (U, V) 的全体为 Nash 谈判集。

对于 2×2 阶双矩阵对策,安全点可由式(11.3)计算。只是注意在计算 V_0 时要将矩阵进行转置。如本例中,

$$U_0 = \frac{1 \times 4 - 0 \times 0}{1 - 0 - 0 + 4} = \frac{4}{5}$$

$$V_0 = \frac{4 \times 1 - 0 \times 0}{4 - 0 - 0 + 1} = \frac{4}{5}$$

易见, Nash 谈判集即线段 AB , 最优得失值 (U^*, V^*) 将在 AB 中产生。这样得到的 (U^*, V^*) 称为 Nash 谈判解。下面给出求 Nash 谈判解的步骤:

①求安全点 (U_0, V_0) ;

②求可行得失值区域,对于 2×2 阶双矩阵对策,即以 (a_{ij}, b_{ij}) ($i=1, 2, j=1, 2$) 为顶点的

四边形;

③求出 Pareto 最优点的区域,在 2×2 阶情形一般位于右上方边界;

④求出 Nash 谈判集;

⑤在 Nash 谈判集中求 $f(U, V) = (U - U_0)(V - V_0)$ 的最大点 (U^*, V^*) , 即 Nash 谈判解。

求 Nash 谈判解的过程很像一个双方谈判协商的过程,称为 Nash 谈判过程。

例 12.3(合作的夫妇之争) 已经求得 Nash 谈判集为线段 AB , 如图 12.8, 过 A, B 的直线表达式为

$$\frac{V-4}{1-4} = \frac{U-1}{4-1}$$

即

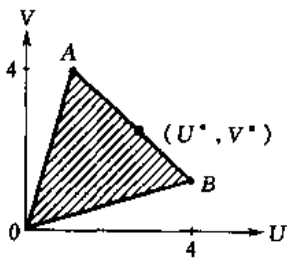


图 12.8

$$V = -U + 5$$

$$\begin{aligned} f(U, V) &= \left(U - \frac{4}{5}\right) \left(V - \frac{4}{5}\right) \\ &= \left(U - \frac{4}{5}\right) \left(-U + 5 - \frac{4}{5}\right) \\ &= -U^2 + 5U - \frac{84}{25} \\ &= -\left(U - \frac{5}{2}\right)^2 + \frac{289}{100} \end{aligned}$$

$$U^* = \frac{5}{2} = 2\frac{1}{2}, V^* = \frac{5}{2} = 2\frac{1}{2}$$

(U^*, V^*) 即 Nash 谈判解。

例 12.4 求双矩阵

$$\text{甲} \begin{bmatrix} (1, 2) & (4, 5) \\ (7, 1) & (3, 0) \end{bmatrix} \quad \text{乙}$$

对策的 Nash 谈判解。

$$\text{解 矩阵 } A = \begin{bmatrix} 1 & 4 \\ 7 & 3 \end{bmatrix}, B^T = \begin{bmatrix} 2 & 1 \\ 5 & 0 \end{bmatrix}$$

$$U_0 = \frac{1 \times 3 - 4 \times 7}{1 - 4 - 7 + 3} = 3\frac{4}{7}, V_0 = \frac{2 \times 0 - 1 \times 5}{2 - 1 - 5 + 0} = 1\frac{1}{4}$$

安全点为 $(U_0, V_0) = \left(3\frac{4}{7}, 1\frac{1}{4}\right)$ 。Nash 谈判集为如

图 12.9 中的线段 AE 。

直线 AD 的方程为

$$\frac{V-5}{1-5} = \frac{U-4}{7-4}$$

即

$$V = -\frac{4}{3}U + \frac{31}{3}$$

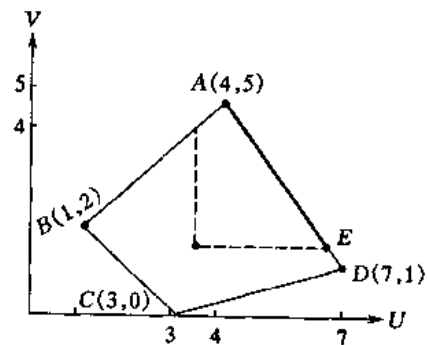


图 12.9

当 $V = V_0 = 1\frac{1}{4}$, $U = 6\frac{3}{16}$, 即 E 点为 $(6\frac{3}{16}, 1\frac{1}{4})$, 在 $4 \leq U \leq 6\frac{3}{16}$ 上求

$$\begin{aligned} f(U) &= \left(U - 3\frac{4}{7}\right) \left(V - 1\frac{1}{4}\right) \\ &= \left(U - 3\frac{4}{7}\right) \left(-\frac{4}{3}U + \frac{109}{12}\right) \\ &= -\frac{4}{3}U^2 + \frac{3489}{252}U - \frac{2725}{84} \end{aligned}$$

的最大点, 令 $f'(U) = 0$, $U^* = \frac{3489}{672} \approx 5.2$, 由于 $f''(U^*) = -\frac{8}{3} < 0$, 故 $U^* = 5.2$ 是极大点。

相应的 $V^* = 3.4$, $(U^*, V^*) = (5.2, 3.4)$ 是 Nash 谈判解。

习 题

12.1 对于下面的双矩阵对策问题, 分别对非合作与合作(用 Nash 谈判过程)两种情形求解。

$$(1) \begin{bmatrix} (1, 3) & (0, 0) \\ (0, 0) & (3, 1) \end{bmatrix}, (2) \begin{bmatrix} (6, 5) & (3, 15) \\ (5, 9) & (20, 5) \end{bmatrix}$$

12.2 求例 12.2 中的囚犯两难推理问题的 Nash 谈判解。

VI 随机运筹技术

第 13 章 排队系统分析

13.1 基本概念

13.1.1 排队系统的组成

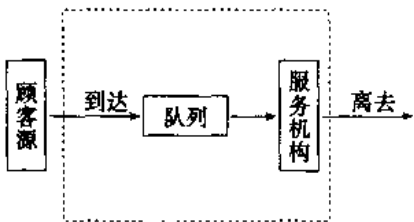


图 13.1

排队是在日常生活中经常遇到的现象。如到商店买东西,到医院看病,均是顾客希望得到某种服务,而当某时刻要求服务的顾客数量超过服务机构的容量时,便出现排队现象。这时,排队等待的顾客与服务机构便构成一个排队系统,如图 13.1 所示。现实世界中存在着形形色色的排队系统,如表 13.1 所示。

表 13.1

到达的顾客	要求服务的内容	服务机构
1. 不能运转的机器	修理	修理技工
2. 修理技工	领取修配零件	发放零件的管理员
3. 病人	就诊	医生
4. 电话呼唤	通话	交换台
5. 文稿	打字	打字员
6. 录入计算机中的文件	打印	打印机
7. 提货单	提取存货	仓库管理员
8. 到达机场上空的飞机	降落	跑道指挥机构
9. 驶入港口的货船	装货或卸货	码头(泊位)
10. 进入水库的上游河水	放水,调整水位	水闸管理员
11. 来到路口的汽车	通过路口	交警或红绿灯
12. 来到车站的乘客	乘车	公交车管理员
13. 进入餐馆的顾客	就餐	餐位的服务员

如果服务设施过少或服务效率太低,便会加剧拥挤,排队成龙。但增加服务设施便会增加服务成本或造成系统空闲,而有些服务设施如机场跑道、港口泊位等一旦建成就不易改动。因此,有必要对排队系统的结构和运行规律加以研究,为排队系统的设计和调控提供依据。

排队系统的基本组成部分主要有输入过程、排队规则、服务机构。

1. 输入过程

输入过程是指顾客到达排队系统的过程,对此我们主要讨论两方面内容:

①顾客源,即顾客总体的数量,可以是有限的,也可以是无限的,河流上游进入水库的水量可以认为是无限的,车间内停机待修的机器显然是有限的;

②顾客到达的规律,主要指顾客到达系统的时间间隔的分布,其类型可以是确定的,如自动生产线上待装配的部件便是按定长时间间隔到达装配点,但更多的情形是随机的。

我们主要讨论到达间隔服从负指数分布的类型。

2. 排队规则

排队规则是指顾客到达系统后排队等候服务的方式和规则。可分为三种类型。

(1) 损失制

指顾客到达系统时,若所有服务设施均被占用,则顾客自动离去,不再回来。如电话系统,当电话呼唤遇到占线则会挂机离去。这实际上是一种不允许排队的规则,系统将“损失”掉许多顾客。

(2) 等待制

指顾客到达系统时,若所有服务设施均被占用,则留下来排队等待,直至被服务完毕后离去。等待制的排队规则又可按顾客被服务的次序分为以下几种。

1) 先到先服务(FCFS) 即按到达次序接受服务,这是最通常的情形。

2) 后到先服务(LCFS) 即后来的反而先被服务。如乘用电梯的顾客常是后人的先出,仓库中存放的厚钢板也是如此。在情报系统中,最后到达的信息往往是最有价值的,常首先加以分析和利用。

3) 具有优先权的服务(PS) 即服务机构根据顾客的优先权进行服务,优先权高的顾客先接受服务。如医院对于病情危重的患者将给予优先治疗。

本书主要讨论先到先服务的最一般情形。

(3) 混合制

这是损失制和等待制的混合。这种排队规则既允许排队又不允许队列无限长,主要包括系统容量有限制和等待时间有限制两种情况。系统容量有限制是指系统中只能容纳或服务 N 个顾客,那么第 N 个及其以前到达的顾客是等待制,而第 $N+1$ 个及以后到达的顾客便是损失制。等待时间有限制是指顾客在系统中的等待时间不能超过某一固定长度 T 。当等待时间超过 T 时,顾客自动离去。如易损电子元件的有效期为 T ,若其在库存待用 T 时间后仍未被使用,便失效不再会被使用。

3. 服务机构

对服务机构的研究内容主要有服务设施(称为服务台)的数量和服务的规律。

(1) 服务台数量

服务台的数量 C 可分为等于 1 和大于 1 两种情形。 $C=1$ 表示服务机构中只有单个服务设施,每次只能为 1 个顾客提供服务; $C>1$ 表示服务机构中有多个服务设施,可同时为 C 个顾客提供服务。对于多个服务台的系统,服务台的排列方式可以是平行并列也可以是依次串列,顾客的队列形式也可以是单队或并列多队。

(2) 服务规律

这主要指服务台对顾客的服务时间的分布。与到达时间的分布一样,服务时间可以是确

定的,如汽车自动冲洗机的冲洗时间为定长,但更多的情形是随机的。本书主要讨论服务时间服从负指数分布的情形。

13.1.2 排队模型的表示

根据输入过程、排队规则和服务机构的不同情况对排队系统进行描述和分类,可以给出很多排队模型。为了方便对众多模型的描述,康道尔(D. G. Kendall)在 1953 年提出一种依据排队系统的三个最基本特征(顾客到达时间间隔的分布、服务时间的分布和服务台数)对排队模型进行分类的表示方法,称为 Kendall 记号。其一般形式为

$$X/Y/Z$$

其中 X 表示顾客到达时间间隔的分布; Y 表示服务时间的分布; Z 表示并列的服务台的个数。

顾客到达时间间隔的分布与服务时间的分布,除了主要讨论的负指数分布之外,还有其他一些分布类型,其名称及简写符号列举如下:

M 表示负指数分布

E_k 表示 k 阶爱尔朗分布

D 表示定长分布

G 表示一般分布

例如, $M/M/1$ 表示顾客到达的间隔时间服从负指数分布、服务台的服务时间也服从负指数分布和单服务台的模型; $D/M/C$ 表示确定的到达间隔、服务时间均为负指数分布和 C 个平行服务台(但顾客是一队)的模型。

后来,在 1971 年的一次关于排队论符号标准化会议上决定,将 Kendall 记号扩充为

$$X/Y/Z/A/B/C$$

的形式,其中前三项意义不变,而

A 表示系统容量限制 N

B 表示顾客源的数目 m

C 表示服务规则,如先到先服务 FCFS,后到先服务 LCFS 等

并约定,如略去后三项,即指 $X/Y/Z/\infty/\infty/FCFS$ 的情形。在本书中,因只讨论先到先服务 FCFS 的情形,所以略去第六项。

13.1.3 排队问题的求解

研究排队系统的目的是通过了解系统运行的状况,对系统进行调整和控制,使系统处于最优运行状态。而描述系统运行状况的客观数量指标主要是系统中由于排队和被服务而滞留的顾客数量,以及顾客为等待服务而必须在系统中消耗的时间。因此,排队论中对排队问题的求解,一个基本内容就是研究和计算这些描述系统运行状态的数量指标,简称系统运行指标。这些数量指标一般来说都是随机变量,并且和系统运行的时间 t 有关。这里主要研究当 $t \rightarrow \infty$ 时的稳态情形。这时系统处于平衡状态,数量指标的分布等与时间无关,这时求得的结果称为系统处于统计平衡下的解。主要研究下述两类指标的概率分布和数字特征,如平均值。

1. 队长和排队长

队长是指系统中的顾客数(排队等待的顾客数与正在接受服务的顾客数之和),其概率分布称为系统的状态概率,记为 P_n ,表示系统中有 n 个顾客的概率。其均值记为 L_s 。

排队长又称队列长,是指系统中正在排队等待服务的顾客数。其均值记为 L_q 。

2. 逗留时间和等待时间

逗留时间是指一个顾客在系统中的停留时间(排队等待时间与被服务时间之和),记为 W ,其均值记为 W_s 。

等待时间又称虚等时间,是指一个顾客在系统中排队等待服务的时间,其均值记为 W_q 。

上述两类数量指标是排队问题求解的最基本内容。除此之外,还有一些数量指标,如忙期(系统连续工作的时间)和闲期(系统连续空闲的时间),服务强度(系统忙着的概率)和顾客损失率(在损失制或系统容量有限的情况下,由于顾客被拒绝而使系统受到的损失)等。而研究这些数量指标的基础是顾客到达(输入)系统的规律以及系统对顾客服务的规律。

13.2 到达与服务的规律

本节研究顾客到达系统的规律以及系统对顾客服务的规律。这里的“规律”主要是指随机变量的概率分布。

13.2.1 到达的规律

可以从两个角度描述顾客到达系统的规律:一是在一定时间里(如单位时间)到达系统的顾客数量;二是顾客相继到达系统的间隔时间。

在概率论中已知,若随机变量 N 的概率分布为

$$P\{N=n\} = \frac{\lambda^n}{n!} e^{-\lambda} \quad (n=0,1,\dots;\lambda>0)$$

则称 N 服从泊松(Poisson)分布。还知道在实际中很多随机现象都近似泊松分布,例如在单位时间(如1分钟)内通过某路口的车辆数、到达某交换台的电话呼唤数,到达某商店的顾客数等等。总之,在1单位时间内到达某排队系统的顾客数 N 是服从泊松分布的随机变量。那么,在 t 单位时间内到达该排队系统的顾客数若记为 $N(t)$,则其概率分布表达式应为

$$P\{N(t)=n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (n=0,1,\dots;\lambda>0;t>0)$$

称 $N(t)$ 为泊松过程,又称泊松流。

1. 泊松流

定义 13.1 称满足下列三条件的输入为泊松流(也称泊松过程或最简单流):无后效性,即在不相交的时间区间内到达的顾客数是相互独立的;平稳性,即在 $[t, t+\Delta t)$ 时段内有1个顾客的概率与 t 无关,只与 Δt 有关(均匀的),也即 $P\{N[t, t+\Delta t)=1\} \approx \lambda \Delta t$ (记为 $P_1(\Delta t) \approx \lambda \Delta t$);普通性,也称稀有性,即在瞬时内只可能有1个顾客到达,也即对充分小的 Δt , $\sum_{n=2}^{\infty} P_n(\Delta t) = 0$ 。

可以证明,如果一个排队系统的顾客输入是满足上述条件的泊松流,那么在长为 t 的时段内到达 n 个顾客的概率

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (n=0,1,\dots;\lambda>0;t>0)$$

下面分析参数 λ 的含义。由泊松分布的数字特征可知, 均值 $E\{N(t)\} = \lambda t$, 说明 λ 为单位时间内的平均到达顾客数。又由 $P_1(\Delta t) = \lambda \Delta t$ 可知, λ 为单位时间内有 1 个顾客到达的概率。实际上, 两种解释是一致的。例如, 某排队系统平均每 5 分钟来 3 名顾客, 那么平均每 1 分钟来 $\frac{3}{5}$ 名顾客, 也可以说平均每分钟有 1 个顾客到达的概率是 $\frac{3}{5}$ 。

以上从顾客到达数这一角度, 就实际中最常见的泊松流输入讨论了顾客到达的规律。下面将从顾客到达间隔时间的角度, 就最常见的负指数间隔进行讨论, 并讨论这两种角度之间的关系。

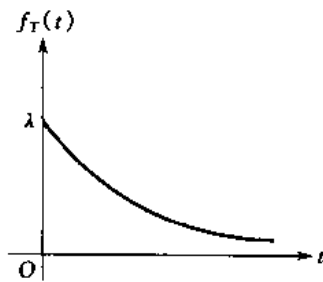


图 13.2

2. 负指数分布

定义 13.2 若随机变量 T 的分布密度为

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

则称 T 服从参数为 λ 的负指数分布 ($\lambda > 0$)。它是连续型分布, 其密度函数如图 13.2 所示。

负指数分布的分布函数为

$$F_T(t) = P(T \leq t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

其均值 $E(T) = \frac{1}{\lambda}$, 方差 $D(T) = \frac{1}{\lambda^2}$ 。

负指数分布有下列性质。

①若 T 服从参数为 λ 的负指数分布, 则有

$$P\{T > t_0 + t | T > t_0\} = P\{T > t\}.$$

证 由条件概率公式得

$$\begin{aligned} P\{T > t_0 + t | T > t_0\} &= P\{(T > t_0 + t) \cap (T > t_0)\} / P\{T > t_0\} \\ &= P\{T > t_0 + t\} / P\{T > t_0\} \\ &= e^{-\lambda(t_0 + t)} / e^{-\lambda t_0} \\ &= e^{-\lambda t} = P\{T > t\} \end{aligned}$$

这个性质称为无记忆性。它表明, 在已知 $T > t_0$ 条件下估计 $T > t_0 + t$ 的概率与不考虑 $T > t_0$ 这一条件而直接估计 $T > t$ 的概率是相同的, 也可以说 T 对于它已大于 t_0 这一点“无记忆”。如果 T 表示一种电子元件的使用寿命, 那么当元件已经用了 t_0 年后再估计它还能用 t 年的概率, 与它全新时估计能用 t 年的概率一样, 即它对已使用了的 t_0 年无记忆, 是一种极其耐磨损的高精度元件。

在排队系统中, 如果顾客到达的间隔时间 T 是服从负指数分布的, 那么在已知间隔超过 t_0 的条件下估计间隔还要跨越 t 的概率与直接估计间隔超过 t 的概率相同, 也可以说顾客到达的间隔时间是纯随机的。由于无记忆性又称马尔可夫 (A. A. Марков) 性, 故负指数分布的表示简记为 M 。

②若排队系统的顾客到达数 $N(t)$ 为泊松流, 则顾客的到达间隔 T 服从负指数分布, 且参数 λ 相同。

证 因为到达为泊松流, 所以 t 时段内没有顾客到来的概率

$$P_0(t) = \frac{\lambda^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

从而 t 时段内有顾客到来(也即间隔不超过 t) 的概率

$$P(T \leq t) = 1 - e^{-\lambda t}$$

即 T 的分布函数 $F(t) = 1 - e^{-\lambda t}$, 说明 T 服从参数为 λ 的负指数分布。

可以证明, 这一性质反之也成立, 故有“到达数为泊松流的充要条件是到达间隔服从负指数分布, 且参数相同”。前面已从泊松流的均值角度解释了参数 λ 的含义, 即 λ 表示单位时间内到达系统的平均顾客数, 现在再从负指数分布的均值的角度来解释 λ 的含义。由于 $E(T) = \frac{1}{\lambda}$, 说明 λ 的倒数表示平均每两位顾客到达的间隔时间。显然这两种角度的解释是等价的, 例如若平均间隔 $\frac{1}{5}$ 小时来 1 位顾客, 则平均每小时到达的顾客数是 5 人。

13.2.2 服务的规律

描述排队系统的服务规律主要是采用系统对顾客服务时间 v 的分布。常见的分布有定长分布、负指数分布、爱尔朗分布等。

这里主要讨论服务时间 v 服从负指数分布的情形。设分布参数为 μ 。由于 $E(v) = \frac{1}{\mu}$, 所以对 μ 的含义的解释是: 系统对每位顾客的平均服务时间为 $\frac{1}{\mu}$, 或者说系统平均每单位时间服务完 μ 个顾客, 即系统的平均服务率是 μ 。

当排队系统的服务机构由 k 个串联的服务台组成时, 每个服务台的服务时间相互独立, 且服从相同的负指数分布(参数设为 $k\mu$), 那么一顾客走完这 k 个服务台总共需要的服务时间就服从 k 阶爱尔朗分布。

定义 13.3 若随机变量 v_1, \dots, v_k 独立且同服从参数为 $k\mu$ 的负指数分布, 则 $T = \sum_{i=1}^k v_i$ 服从 k 阶爱尔朗(Erlang)分布, 密度为

$$f_T(t) = \frac{k\mu (k\mu t)^{k-1}}{(k-1)!} e^{-k\mu t} \quad (t \geq 0)$$

爱尔朗分布的期望与方差分别是

$$E(T) = \frac{1}{\mu}, D(T) = \frac{1}{k\mu^2}$$

爱尔朗分布是随 k 而变化的一族分布。当 $k=1$ 时化为负指数分布; 而当 $k \rightarrow +\infty$ 时, 由于 $D(T) \rightarrow 0$, 故化为确定型分布; 当 $1 < k < +\infty$ 时, 它是介于纯随机和完全确定之间的中间型分布, 能对现实世界提供更加广泛的适应性。爱尔朗分布族密度函数如图 13.3 所示。

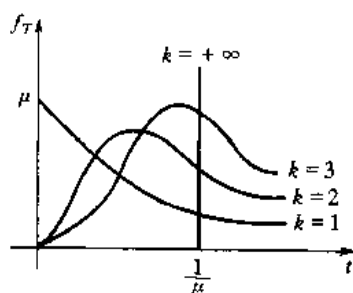


图 13.3

13.2.3 分布律的检验

对于一个实际的排队系统, 如何检验它的顾客到达是否形成泊松流, 或者它的服务机构的服务时间是否服从负指数分布, 是一个重要的现实问题。可以根据数理统计中检验分布律的

方法(如 χ^2 检验法)解决。该方法步骤如下:

①提出假设 H_0 , 总体分布 $F(x) = F_0(x)$ (其中 $F_0(x)$ 为某已知分布函数);

②在总体 X 中抽取样本点 x_1, \dots, x_n ;

③将 x 轴分为 m 个不相交的小区间 $(-\infty, t_1], (t_1, t_2], \dots, (t_{m-1}, +\infty)$, 记落入第 i 个小区间 $(t_{i-1}, t_i]$ 中的样本点个数为 f_i ;

④计算皮尔逊统计量 $V = \sum_{i=1}^m \frac{(f_i - np_i)^2}{np_i}$, 其中 $p_i = F_0(t_i) - F_0(t_{i-1})$;

⑤由 χ^2 表查得临界值 $\chi_{\alpha}^2(m-r-1)$, 其中 r 为 $F_0(x)$ 中未知参数的个数;

⑥若 $V < \chi_{\alpha}^2(m-r-1)$, 则肯定 H_0 , 否则否定 H_0 。

例 13.1 某排队系统为了检验每周的顾客到达是否服从泊松分布, 连续观察了 52 周的到达数量, 数据如表 13.2 所示。

表 13.2

周	顾客到达数												
1~13	5	5	2	1	3	3	3	3	8	6	5	3	6
14~26	4	6	5	3	2	6	2	5	3	2	8	2	2
27~39	6	3	6	2	5	4	6	4	5	3	4	2	2
40~52	6	6	3	7	5	0	4	4	10	1	3	4	0

根据顾客到达情况, 将观测值的范围分为 8 个组: 0~1 为一组, 2、3、4、5、6、7 各为一组, 8 以上为一组。为了查理论分布值, 计算 52 个样本值的均值 $\bar{x} = 4$ 。然后由 $\lambda = 4$ 的泊松分布查得各组的概率值 p_i , 如 $p_1 = P(0) + P(1) = 0.0183 + 0.0733 = 0.0916$, $p_2 = P(2) = 0.1465$, 再计算 $F_i = np_i = 52p_i$, $f_i - F_i$ 和 $(f_i - F_i)^2 / F_i$, 计算结果列于表 13.3。

表 13.3

组	到达数	观测频数	理论频率	理论频数	残差	相对残差平方
i	x	f_i	p_i	F_i	$f_i - F_i$	$(f_i - F_i)^2 / F_i$
1	0~1	4	.0916	4.763	-.763	.122
2	2	9	.1465	7.613	1.382	.251
3	3	11	.1954	10.161	.839	.069
4	4	7	.1954	10.161	-3.161	.983
5	5	8	.1563	8.128	-.128	.002
6	6	9	.1042	5.418	3.582	2.363
7	7	1	.0595	3.094	-2.094	1.417
8	8 以上	3	.0511	2.657	.343	.044
	总计	52	1.0000	52.000	0.000	5.256

由于 $F_0(x)$ 中有一个未知参数 λ , 故 $r = 1$ 。取检验的水平 $\alpha = 0.05$, 查 χ^2 表得临界值 $\chi_{0.05}^2(8-1-1) = 12.592$ 。由于 $V = 5.256 < 12.592$, 故应肯定 H_0 , 即可认为顾客的到达是服从泊松分布, 其参数估计值 $\lambda = 4$ 。

13.3 M/M/1 排队模型

本节研究具有泊松输入和负指数服务的单服务台排队系统, 分三种情况讨论排队模型。

13.3.1 标准的 M/M/1 模型(M/M/1/ ∞/∞)

标准的 M/M/1 模型是针对适合下列条件的排队系统:

①输入过程——顾客源是无限的,顾客单个到来,到达的间隔时间服从负指数分布,设其参数为 λ ;

②排队规则——单队,对队长没有限制,先到先服务;

③服务机构——单服务台,服务时间服从负指数分布,设参数为 μ 。

1. 系统的状态概率

系统中的状态是指系统中的顾客数量,记为 n 。对于标准的 M/M/1 系统, $n=0,1,\dots$, 系统的状态概率记为 P_n , 表示系统中有 n 个顾客的概率。对于稳态的情形,首先可将系统状态间的转移关系用图 13.4 表示。图中的 \textcircled{n} 表示状态 n ($n=0,1,\dots$)。由状态 $n-1$ 转移到状态 n , 是由于又来了一位顾客,而顾客的到达率为 λ , 故在图中表示相邻状态增加的箭线上标上 λ , 转移的概率是 λP_{n-1} , 表示原来是状态 $n-1$ 且又来了一位。由状态 $n+1$ 转移到状态 n , 是由于服务完了一位顾客,而系统的服务率为 μ , 故在图中表示相邻状态减少的箭线上标上 μ , 转移的概率是 μP_{n+1} , 表示原来是状态 $n+1$ 且又走了一位。

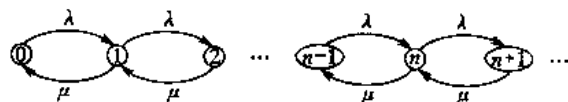


图 13.4

对于状态 n ($n=1,2,\dots$), 由于系统满足统计平衡, 即流入应等于流出, 故有

$$\lambda P_{n-1} + \mu P_{n+1} = \lambda P_n + \mu P_n$$

对于状态 0, 也应有

$$\lambda P_0 = \mu P_1$$

将二式联立, 即系统状态的平衡方程组

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n \quad (n=1,2,\dots) \end{cases}$$

下面求解平衡方程组, 以得到状态概率。首先由第一个方程解得

$$P_1 = \frac{\lambda}{\mu} P_0$$

在第二个方程中令 $n=1$, 可得到

$$P_2 = \frac{\lambda}{\mu} P_1 = \left(\frac{\lambda}{\mu}\right)^2 P_0$$

再分别令 $n=2,3,\dots$, 可得到

$$P_n = \frac{\lambda}{\mu} P_{n-1} = \left(\frac{\lambda}{\mu}\right)^n P_0$$

记 $\rho = \frac{\lambda}{\mu}$, 并设 $\rho < 1$ (否则队列将排至无限远), 由概率性质, 有

$$\sum_{n=0}^{\infty} P_n = 1$$

将 P_n 的表达式代入,有

$$P_0 \sum_{n=0}^{\infty} \rho^n = P_0 \frac{1}{1-\rho} = 1$$

可得状态概率

$$\begin{cases} P_0 = 1 - \rho \\ P_n = \rho^n P_0 = (1 - \rho) \rho^n \quad (n = 1, 2, \dots) \end{cases}$$

式中的 ρ 有明确的实际意义: 由 $\rho = \frac{\lambda}{\mu}$ 说明, ρ 是平均到达率与平均服务率之比, 由 $\rho = \frac{\frac{1}{\mu}}{\frac{1}{\lambda}}$ 说

明 ρ 是平均服务时间与平均到达间隔时间之比, 反映了系统的服务强度。而由 $\rho = 1 - P_0$ 说明 ρ 是系统忙着的概率, 反映了系统的繁忙程度和利用率。因此, 称 ρ 为服务强度或服务机构的利用率。

2. 系统的运行指标

基于状态概率, 可计算出系统的运行指标。

(1) 平均队长和排队长

队长的平均值 L_s 表示系统中的平均顾客数, 表达式为

$$\begin{aligned} L_s &= \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n \rho^n (1 - \rho) \\ &= (\rho + 2\rho^2 + 3\rho^3 + \dots) - (\rho^2 + 2\rho^3 + 3\rho^4 + \dots) \\ &= \rho + \rho^2 + \rho^3 + \dots = \frac{\rho}{1 - \rho}, 0 < \rho < 1 \end{aligned}$$

或

$$L_s = \frac{\lambda}{\mu - \lambda}$$

排队长又称队列长, 其平均值 L_q 表示系统中排队等待的平均顾客数。由于系统有一个服务台, 故其表达式为

$$L_q = \sum_{n=1}^{\infty} (n-1) P_n = \sum_{n=0}^{\infty} n P_n - \sum_{n=1}^{\infty} P_n = L_s - \rho = \frac{\rho^2}{1 - \rho}$$

注意这里的队长与排队长之差直观上看似乎应为 1, 但上面的计算说明 $L_s - L_q = \rho < 1$, 这是因为它们的差在系统忙时是 1, 而在系统闲时是 0, 故其均值为 $0 \cdot P_0 + 1 \cdot (1 - P_0) = \rho$ 。

(2) 平均逗留时间和等待时间

首先可证, 顾客在系统中的逗留时间 W 服从参数为 $\mu - \lambda$ 的负指数分布。

事实上, 设一顾客到达时系统中已有 n 个顾客, 则该顾客在系统中的逗留时间 $W_n = v'_1 + v_2 + \dots + v_n + v_{n+1}$, 其中 v'_1 为正被服务的第 1 位顾客还需要的服务时间, $v_i (i = 2, \dots, n)$ 为第 i 位顾客的被服务时间, v_{n+1} 为该顾客的被服务时间。

因为 $v_i (i = 2, \dots, n+1)$ 独立且同服从参数为 μ 的负指数分布, 而 v'_1 由无记忆性也服从参数为 μ 的负指数分布, 所以 W_n 服从 $n+1$ 阶爱尔朗分布, 即

$$f_{W_n}(t) = \frac{\mu(\mu t)^n}{n!} e^{-\mu t}$$

从而 W 的分布函数

$$\begin{aligned} F_W(t) &= P(W \leq t) = \sum_{n=0}^{\infty} P_n P\{W \leq t | n\} \\ &= \sum_{n=0}^{\infty} \rho^n (1-\rho) \int_0^t \frac{\mu(\mu t)^n}{n!} e^{-\mu t} dt \\ &= 1 - e^{-(\mu-\lambda)t} \end{aligned}$$

即 W 服从参数为 $\mu - \lambda$ 的负指数分布, 故其均值

$$W_s = \frac{1}{\mu - \lambda}$$

而平均等待时间应等于平均逗留时间与平均服务时间之差, 故

$$W_q = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}$$

(3) 运行指标之间的关系

先将上述 4 个系统运行指标归纳如下:

$$\begin{aligned} L_s &= \frac{\lambda}{\mu - \lambda} & L_q &= \frac{\rho\lambda}{\mu - \lambda} \\ W_s &= \frac{1}{\mu - \lambda} & W_q &= \frac{\rho}{\mu - \lambda} \end{aligned}$$

易见它们之间的关系如下:

$$\begin{aligned} L_s &= \lambda W_s & L_q &= \lambda W_q \\ L_s - L_q &= \frac{\lambda}{\mu} & W_s - W_q &= \frac{1}{\mu} \end{aligned}$$

这一组关系式称为里特(Little)公式。里特公式对于其他一些排队模型也成立, 只是式中的 λ 一般应为平均的顾客实际进入系统率, 也称有效到达率, 记为 λ_e 。在本模型中因为到来的顾客均能进入系统, 故 $\lambda_e = \lambda$ 。

例 13.2 某修理店只有一个修理工, 来修理的顾客到达过程为泊松流, 平均每小时 4 人。修理时间服从负指数分布, 平均需要 6 min。试求①修理店空闲的概率; ②店内有 3 个顾客的概率; ③店内至少有 1 个顾客的概率; ④在店内的平均顾客数; ⑤等待服务的平均顾客数; ⑥每位顾客在店内的平均逗留时间; ⑦每位顾客平均等待修理的时间; ⑧顾客在店内需消耗 15 min 以上的概率。

解 本例可看做一个标准的 $M/M/1$ 模型, $\lambda = 4$ 人/h, $\mu = \frac{1}{6}$ 人/min = 10 人/h, $\rho = \frac{\lambda}{\mu} = \frac{2}{5}$ 。

①修理店空闲的概率

$$P_0 = 1 - \rho = 1 - \frac{2}{5} = \frac{3}{5} = 0.6$$

②店内有 3 个顾客的概率

$$P_3 = \rho^3 (1 - \rho) = \left(\frac{2}{5}\right)^3 \left(1 - \frac{2}{5}\right) = 0.038$$

③店内至少有 1 个顾客的概率

$$1 - P_0 = \rho = \frac{2}{5} = 0.4$$

④店内顾客平均数

$$L_s = \frac{\rho}{1 - \rho} = \frac{\frac{2}{5}}{1 - \frac{2}{5}} = \frac{2}{3} = 0.67 \text{ 人/h}$$

⑤等待服务的平均顾客数

$$L_q = L_s - \rho = \frac{2}{3} - \frac{2}{5} = \frac{4}{15} = 0.267 \text{ 人/h}$$

⑥顾客在店内的平均逗留时间

$$W_s = \frac{L_s}{\lambda} = \frac{2}{3} \times \frac{1}{4} = \frac{1}{6} \text{ h/人} = 10 \text{ min/人}$$

⑦顾客平均等待修理时间

$$W_q = W_s - \frac{1}{\mu} = \frac{1}{6} - \frac{1}{10} = \frac{1}{15} \text{ h/人} = 4 \text{ min/人}$$

⑧顾客在店内需消耗 15 分钟以上的概率

$$\begin{aligned} P\left(W \geq \frac{1}{4}\right) &= 1 - P\left(W < \frac{1}{4}\right) \\ &= 1 - e^{-(10-4)\frac{1}{4}} \\ &= e^{-1.5} = 0.223 \end{aligned}$$

例 13.3 考虑一个铁路列车编组站, 设待编列车到达时间间隔服从负指数分布, 平均每小时到达 2 列; 编组站的编组时间也服从负指数分布, 平均每 20 分钟可编一组。该编组站内共有 2 股车道, 当均被占用时, 再来的列车只能停在站外或前方站。①求该编组系统中列车的平均数, 每一列车的平均停留时间和等待编组的列车平均数。②如果列车因站内的 2 股车道均被占用而停在站外或前方站时, 每列车的费用为每小时 a 元, 求每天由此而造成的平均费用损失。

解 本例可看做一个标准的 M/M/1 系统, $\lambda = 2 \text{ 列/h}$, $\mu = \frac{1}{20} \text{ 列/h} = 3 \text{ 列/min}$, $\rho = \frac{\lambda}{\mu} = \frac{2}{3}$

①编组系统中列车的平均数

$$L_s = \frac{\rho}{1 - \rho} = \frac{\frac{2}{3}}{1 - \frac{2}{3}} = 2 \text{ 列/h}$$

列车的平均停留时间

$$W_s = \frac{L_s}{\lambda} = \frac{2}{2} = 1 \text{ h/列}$$

等待编组的列车平均数

$$L_q = L_s - \rho = 2 - \frac{2}{3} = \frac{4}{3} \text{ 列/h}$$

②每天由于列车站外等待而造成的平均费用为

$$24\lambda W_s (1 - P_0 - P_1 - P_2) a = 24 \times 2 \times \left(\frac{2}{3}\right)^3 a = 14.2a \text{ 元}$$

13.3.2 系统容量有限的 M/M/1 模型(M/M/1/N/∞)

本模型与标准的 M/M/1 模型的主要区别有二:一是系统的状态只有 $N+1$ 种($n=0,1,\dots,N$);二是顾客的实际进入系统率当 $n < N$ 时为 λ ,而当 $n = N$ 时为 0,故平均值(即有效到达率)

$$\lambda_e = \lambda(1 - P_N) + 0 \cdot P_N = \lambda(1 - P_N)$$

1. 系统的状态概率

首先仍可画出系统的状态转移图,如图 13.5 所示。由统计平衡,列出系统状态的平衡方程组

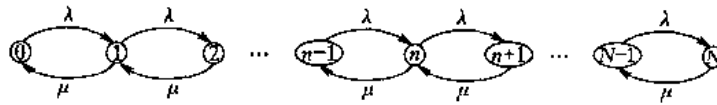


图 13.5

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu) P_n \quad (n=1, \dots, N-1) \\ \lambda P_{N-1} = \mu P_N \end{cases}$$

求解平衡方程组,首先可得 $P_n = \frac{\lambda}{\mu} P_{n-1} = \left(\frac{\lambda}{\mu}\right)^n P_0$,记 $\rho = \frac{\lambda}{\mu}$ 。当 $\rho \neq 1$,有

$$\begin{aligned} \sum_{n=0}^N P_n &= P_0(1 + \rho + \dots + \rho^N) \\ &= P_0 \frac{1 - \rho^{N+1}}{1 - \rho} = 1 \end{aligned}$$

解得
$$\begin{cases} P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} \\ P_n = \rho^n P_0 \quad (n=1, \dots, N) \end{cases}$$

当 $\rho = 1$,由平衡方程组可得 $P_0 = P_1 = \dots = P_N = \frac{1}{N+1}$ 。

状态概率中的 P_N 表示系统满员的概率,因此也可称为系统对顾客的拒绝率或顾客的损失率。

2. 系统的运行指标

先求出系统的平均队长,当 $\rho \neq 1$ 时有

$$\begin{aligned} L_s &= \sum_{n=0}^N n P_n = P_0 \sum_{n=0}^N n \rho^n = P_0 \rho \frac{d}{d\rho} \left(\sum_{n=0}^N \rho^n \right) \\ &= P_0 \rho \frac{d}{d\rho} \left(\frac{1 - \rho^{N+1}}{1 - \rho} \right) = P_0 \rho \frac{1 - (N+1)\rho^N + N\rho^{N+1}}{(1 - \rho)^2} \\ &= \frac{\rho[1 - (N+1)\rho^N + N\rho^{N+1}]}{(1 - \rho^{N+1})(1 - \rho)} \\ &= \frac{\rho}{1 - \rho} - \frac{(N+1)\rho^{N+1}}{1 - \rho^{N+1}} \end{aligned}$$

当 $\rho = 1$ 时有

$$L_s = \frac{1}{N+1}(1+2+\cdots+N) = \frac{N}{2}$$

再由里特公式, 求出其他三个指标。当 $\rho \neq 1$ 时, 平均排队长

$$L_q = L_s - (1 - P_0) = L_s - \frac{\rho(1+N\rho^N)}{1-\rho^{N+1}} = \frac{\rho}{1-\rho} - \frac{\rho(1+\rho^N)}{1-\rho^{N+1}}$$

当 $\rho = 1$ 时, 平均排队长

$$L_q = L_s - (1 - P_0) = \frac{N}{2} - \left(1 - \frac{1}{N+1}\right) = \frac{N(N-1)}{2(N+1)}$$

平均逗留时间 $W_s = L_s / \lambda_e$, 平均等待时间 $W_q = L_q / \lambda_e$ 或 $W_q = W_s - \frac{1}{\mu}$ 。

这里的有效到达率 $\lambda_e = \lambda(1 - P_N)$ 的直观解释是, 顾客到达且系统未滿而实际进入系统率。由统计平衡知, 它应等于实际服务率 $\mu(1 - P_0)$ 。它表示系统不空且以 μ 的速率服务着。

$\frac{\lambda_e}{\mu}$ 称为本系统的服务强度。

将 M/M/1/N/ ∞ 排队模型的运行指标归纳如下:

$$L_s = \begin{cases} \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}, & \rho \neq 1 \\ \frac{N}{2}, & \rho = 1 \end{cases}$$

$$L_q = L_s - (1 - P_0) = \begin{cases} \frac{\rho}{1-\rho} - \frac{\rho(1+N\rho^N)}{1-\rho^{N+1}}, & \rho \neq 1 \\ \frac{N(N-1)}{2(N+1)}, & \rho = 1 \end{cases}$$

$$W_s = \frac{L_s}{\lambda_e} = \frac{L_s}{\lambda(1-P_N)}$$

$$W_q = \frac{L_q}{\lambda_e} = \frac{L_q}{\lambda(1-P_N)}$$

例 13.4 某修理站只有一个修理工, 且站内最多只能停放 3 台待修的机器。设待修机器按泊松流到达修理站, 平均每分钟到达 1 台; 修理时间服从负指数分布, 平均每 1.25 min 可修理 1 台, 试求: ①空闲率; ②顾客损失率; ③顾客的有效到达率; ④平均队长; ⑤平均排队长; ⑥平均逗留时间; ⑦平均等待时间。

解 本例可看做一个 M/M/1/4/ ∞ 排队系统, $\lambda = 1$ 台/min, $\mu = \frac{1}{1.25} = 0.8$ 台/min, $\rho = \frac{\lambda}{\mu} = \frac{1}{0.8} = 1.25$ 。

$$\text{①空闲率 } P_0 = \frac{1-\rho}{1-\rho^{4+1}} = \frac{1-1.25}{1-1.25^5} = 0.122$$

$$\text{②顾客损失率 } P_4 = \rho^4 P_0 = 1.25^4 \times 0.122 = 0.298$$

$$\text{③有效到达率 } \lambda_e = \lambda(1 - P_4) = 1 \times (1 - 0.298) = 0.702$$

$$\text{④平均队长 } L_s = \frac{\rho}{1-\rho} - \frac{(4+1)\rho^5}{1-\rho^5} = \frac{1.25}{1-1.25} - \frac{5 \times 1.25^5}{1-1.25^5} = 2.44 \text{ 台}$$

$$\text{⑤平均排队长 } L_q = L_s - (1 - P_0) = 2.44 - (1 - 0.122) = 1.56 \text{ 台}$$

$$\textcircled{6} \text{平均逗留时间 } W_s = \frac{L_s}{\lambda_e} = \frac{2.44}{0.702} = 3.48 \text{ min}$$

$$\textcircled{7} \text{平均等待时间 } W_q = \frac{L_q}{\lambda_e} = \frac{1.56}{0.702} = 2.23 \text{ min}$$

例 13.5 某工厂有一个半成品加工操作间,内设一个半成品加工操作台和可存放 3 个待加工半成品的场地。已知半成品按平均每天 3 个的泊松过程到达该操作间,而完成该半成品加工的必要时间服从平均每个需 $\frac{1}{4}$ 天的负指数分布。若半成品到达操作间时操作间内已没有场地存放,则要运往其他地方。①需运往其他地方的半成品占到达操作间的半成品总数的比例是多少? ②现工厂为扩大生产能力正在另建一新厂,厂内也需建一个半成品加工操作间,假设半成品到达及加工时间均与原来的相同,每个半成品的存放需要 1 m^2 。若工厂希望能有 95% 的半成品不被运往其他地方,则操作间的存放场地应修建成多大面积?

解 此问题可看做一个系统容量有限的 M/M/1 系统, $\lambda = 3$ 个/天, $\mu = 4$ 个/天, $\rho = \frac{3}{4}$ 。

①需运走的半成品占全部到达的半成品的比例即系统的拒绝率 P_N , 这里 $N = 4$, 故

$$P_4 = \rho^4 P_0 = \left(\frac{3}{4}\right)^4 \frac{1 - \frac{3}{4}}{1 - \left(\frac{3}{4}\right)^5} = 0.316 \times \frac{1 - 0.75}{1 - 0.237} = 0.104$$

②若要求 $P_N = 1 - 0.95 = 0.05$, 即

$$\left(\frac{3}{4}\right)^N \frac{1 - \frac{3}{4}}{1 - \left(\frac{3}{4}\right)^{N+1}} = 0.05$$

$$\left(\frac{3}{4}\right)^N = 0.2 \left(1 - \left(\frac{3}{4}\right)^{N+1}\right)$$

$$5 + \frac{3}{4} = \left(\frac{4}{3}\right)^N$$

$$N = \log_3 5.75 = 6.07 \approx 6$$

故操作间的存放场地面积应修建成 $6 - 1 = 5 \text{ m}^2$ 。

13.3.3 顾客源有限的 M/M/1 模型(M/M/1/ ∞ /m)

本模型与标准的 M/M/1 模型的主要区别有二:一是系统的状态只有 $m + 1$ 种($n = 0, 1, \dots, m$);二是平均到达率 λ 在无限源时表示全体顾客的平均到达率,即单位时间平均到 λ 人,而在有限源时表示每个顾客的平均到达率,即单位时间每人到 λ 次;三是实际进入系统率,当 $n < m$ 时为 $(m - n)\lambda$,而当 $n = m$ 时为 0,平均到达率

$$\begin{aligned} \lambda_e &= E[(m - n)\lambda] = E(m\lambda) - E(n\lambda) \\ &= m\lambda - \lambda E(n) = m\lambda - \lambda L_s \\ &= \lambda(m - L_s) \end{aligned}$$

1. 系统的状态概率

首先画出系统的状态转移图,如图 13.6 所示。

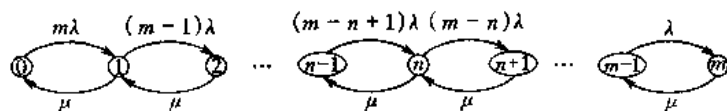


图 13.6

由统计平衡列出系统状态的平衡方程组

$$\begin{cases} m\lambda P_0 = \mu P_1 \\ (m-n+1)\lambda P_{n-1} + \mu P_{n+1} = [(m-n)\lambda + \mu] P_n \quad (n=1, \dots, m-1) \\ \mu P_m = \lambda P_{m-1} \end{cases}$$

求解平衡方程组, 可得系统的状态概率

$$\begin{cases} P_0 = \frac{1}{\sum_{i=0}^m \frac{m!}{(m-i)!} \left(\frac{\lambda}{\mu}\right)^i} \\ P_n = \frac{m!}{(m-n)!} \left(\frac{\lambda}{\mu}\right)^n P_0 \quad (n=1, \dots, m) \end{cases}$$

2. 系统的运行指标

由里特公式得

$$L_s - L_q = \frac{\lambda_e}{\mu} = \frac{\lambda(m - L_s)}{\mu}$$

另一方面, $L_s - L_q = 1 - P_0$, 由此解得

$$L_s = m - \frac{\mu}{\lambda}(1 - P_0)$$

将 L_s 的表达式写成 $\lambda(m - L_s) = \mu(1 - P_0)$, 可以很直观地解释其实际意义: $m - L_s$ 为系统外的顾客数, $\lambda(m - L_s)$ 为系统的实际到达率; $1 - P_0$ 为系统忙着的概率, $\mu(1 - P_0)$ 为系统忙时的服务率, 由统计平衡, 二者应当相等。

由里特公式, 可得另三个指标:

$$L_q = L_s - (1 - P_0)$$

$$W_s = L_s / \lambda_e$$

$$W_q = L_q / \lambda_e$$

在本模型中, 仍将 $\frac{\lambda_e}{\mu}$ 称做服务强度。

例 13.6 某车间有 5 台机器, 每台机器的连续运转时间服从负指数分布, 平均连续运转时间为 15 min。有一个修理工, 每次修理时间服从负指数分布, 平均每次 12 min。求: ①修理工空闲的概率; ②五台机器都出故障的概率; ③出故障机器的平均台数; ④等待修理机器的平均台数; ⑤每台机器的平均停工时间; ⑥每台机器平均等待修理时间。

解 此问题可看做是一个 M/M/1/∞/5 排队系统, $\lambda = \frac{1}{15}$ 台/min, $\mu = \frac{1}{12}$ 台/min, $\frac{\lambda}{\mu} = \frac{12}{15} = 0.8$

①修理工空闲的概率

$$P_0 = \left[\frac{5!}{5!} (0.8)^0 + \frac{5!}{4!} (0.8)^1 + \frac{5!}{3!} (0.8)^2 + \frac{5!}{2!} (0.8)^3 + \frac{5!}{1!} (0.8)^4 + \frac{5!}{0!} (0.8)^5 \right]^{-1}$$

$$= 0.0073$$

②五台机器都出故障的概率

$$P_5 = \frac{5!}{0!} (0.8)^5 P_0 = 0.287$$

③出故障机器的平均台数

$$L_s = 5 - \frac{1}{0.8} (1 - 0.0073) = 3.76 \text{ 台}$$

④等待修理机器的平均台数

$$L_q = 3.76 - (1 - 0.0073) = 2.77 \text{ 台}$$

⑤每台机器平均停工时间

$$W_s = \frac{5}{\frac{1}{12} (1 - 0.0073)} - 15 = 46 \text{ min}$$

⑥每台机器平均待修时间

$$W_q = 46 - 12 = 34 \text{ min}$$

上述结果表明,机器停工时间过长,修理工几乎没有空闲时间,应当提高服务率或增加修理工人。

13.4 M/M/C 排队模型

本节研究具有泊松输入和负指数服务的多服务台排队系统,仍可类似于上节分为三种类型讨论。这里只讨论第一种类型,即系统容量和顾客源均为无限的标准 M/M/C 模型(M/M/C/ ∞/∞)。讨论的前提是系统为单队、并列 C 个服务台,如图 13.7 所示。

标准的 M/M/C 模型与标准的 M/M/1 模型的主要区别一是 M/M/1 模型中的服务率 μ 与系统的状态无关,而 M/M/C 模型的服务率与系统状态有关,当 $n < C$ 为 $n\mu$,当 $n \geq C$ 为 $C\mu$;二是 M/M/1 系统的服务强度 $\rho = \frac{\lambda}{\mu}$,而 M/M/C 系统的服务强度

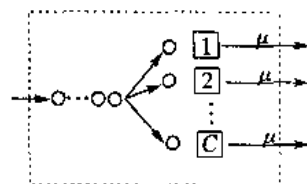


图 13.7

$\rho = \frac{\lambda}{C\mu}$ (仍设 $\rho < 1$), 表示每个服务台单位时间内的平均负荷。

1. 系统的状态概率

首先画出标准的 M/M/C 系统的状态转移图,如图 13.8 所示。

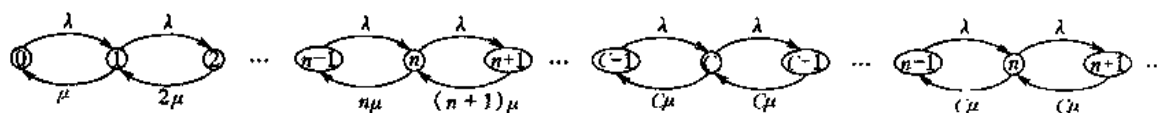


图 13.8

再由统计平衡列出系统状态的平衡方程组

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + (n+1)\mu P_{n+1} = (\lambda + n\mu)P_n \quad (n < C) \\ \lambda P_{n-1} + C\mu P_{n+1} = (\lambda + C\mu)P_n \quad (n \geq C) \end{cases}$$

求解方程组可得

$$P_1 = \frac{\lambda}{\mu} P_0, P_2 = \frac{\lambda}{2\mu} P_1, \dots, P_C = \frac{\lambda}{C\mu} P_{C-1}, P_n = \frac{\lambda}{C\mu} P_{n-1} \quad (n > C)$$

再由 $\sum_{n=1}^{\infty} P_n = 1$ 解出 P_0 , 得到状态概率

$$P_0 = \left[\sum_{k=0}^{C-1} \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k + \frac{1}{C!} \cdot \frac{1}{1-\rho} \cdot \left(\frac{\lambda}{\mu} \right)^C \right]^{-1}$$

$$P_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0 & (n \leq C) \\ \frac{1}{C!} \frac{\left(\frac{\lambda}{\mu} \right)^n}{C^{n-C}} P_0 & (n > C) \end{cases}$$

2. 系统运行指标

平均队长

$$\begin{aligned} L_s &= \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^C n P_n + \sum_{n=C+1}^{\infty} (n-C) P_n + \sum_{n=C+1}^{\infty} C P_n \\ &= \sum_{n=0}^C (n-C) P_n + L_q + C \sum_{n=0}^{\infty} P_n \\ &= L_q + C - \sum_{n=0}^C (C-n) P_n \end{aligned}$$

其中, $C - \sum_{n=0}^C (C-n) P_n$ 的直观意义是平均忙着的服务台数, 而每台平均服务率为 μ , 故系统

平均服务率为 $\mu [C - \sum_{n=0}^C (C-n) P_n]$ 。由统计平衡, 它应等于顾客的平均到达率 λ , 即

$$C - \sum_{n=0}^C (C-n) P_n = \frac{\lambda}{\mu}$$

从而有 $L_s = L_q + \frac{\lambda}{\mu}$

平均排队长

$$\begin{aligned} L_q &= \sum_{n=C+1}^{\infty} (n-C) P_n = \sum_{n=1}^{\infty} n P_{n+C} \\ &= \sum_{n=1}^{\infty} n \cdot \frac{1}{C!} \frac{1}{C^n} (C\rho)^{n+C} P_0 \\ &= \frac{(C\rho)^C}{C!} P_0 \sum_{n=1}^{\infty} n \rho^n = \frac{(C\rho)^C \rho}{C! (1-\rho)^2} P_0 \end{aligned}$$

平均逗留时间

$$W_s = \frac{L_s}{\lambda}$$

平均等待时间

$$W_q = \frac{L_q}{\lambda}$$

将 $M/M/C/\infty/\infty$ 系统运行指标归纳如下:

$$L_s = L_q + \frac{\lambda}{\mu}$$

$$L_q = \frac{(C\rho)^C \rho}{C! (1-\rho)^2} P_0, \text{ 其中 } \rho = \frac{\lambda}{C\mu} < 1$$

$$W_s = \frac{L_s}{\lambda}$$

$$W_q = \frac{L_q}{\lambda}$$

由于 $L_s - L_q = \frac{\lambda}{\mu}$, 故 $W_s - W_q = \frac{L_s - L_q}{\lambda} = \frac{1}{\mu}$, 均与标准的 $M/M/1$ 系统相同, 而与 C 无关。

例 13.7 某售票所有三个窗口, 顾客到达服从泊松流, 平均到达率为每分钟 0.9 人, 售票服务时间服从负指数分布, 平均服务率为每分钟 0.4 人。如果售票所的排队方式是顾客到达后排或一队, 依次向空闲的窗口购票, 求: ①售票所空闲的概率; ②售票所内的平均顾客数和排队等待的平均顾客数; ③顾客为购票需在售票所的平均消耗时间和平均排队等待的时间; ④顾客到达后必须排队等待的概率。

解 这是一个 $M/M/3/\infty/\infty$ 排队系统, $\lambda = 0.9$ 人/min, $\mu = 0.4$ 人/min, $\frac{\lambda}{\mu} = \frac{0.9}{0.4} = 2.25$, $\rho = \frac{\lambda}{3\mu} = \frac{2.25}{3} = 0.75$ 。

①售票所空闲的概率

$$P_0 = \left[\frac{(2.25)^0}{0!} + \frac{(2.25)^1}{1!} + \frac{(2.25)^2}{2!} + \frac{(2.25)^3}{3! (1-2.25/3)} \right]^{-1} = 0.0748$$

②售票所内排队等待的平均顾客数

$$L_q = \frac{(2.25)^3 \times 0.75}{3! (1-0.75)^2} \times 0.0748 = 1.70 \text{ 人}$$

售票所内的平均顾客数

$$L_s = L_q + \frac{\lambda}{\mu} = 1.70 + 2.25 = 3.95 \text{ 人}$$

③顾客的平均消耗时间

$$W_s = \frac{L_s}{\lambda} = \frac{3.95}{0.9} = 4.39 \text{ min}$$

顾客的平均等待时间

$$W_q = \frac{L_q}{\lambda} = \frac{1.70}{0.9} = 1.89 \text{ min}$$

④顾客到达后必须排队等待的概率

$$\begin{aligned} P(n \geq 3) &= 1 - P_0 - P_1 - P_2 \\ &= 1 - 0.0748 - 2.25 \times 0.0748 - \frac{1}{2} \times (2.25)^2 \times 0.0748 \\ &= 0.57 \end{aligned}$$

在本例中,如果顾客的排队方式变为到达售票所后可到任一窗口排队,且入队后不再换队,即可形成3个并行的队列。这时,原来的 $M/M/3/\infty/\infty$ 系统实际上变成了由3个 $M/M/1/\infty/\infty$ 子系统组成的排队系统;且每个子系统的平均到达率为

$$\lambda_1 = \lambda_2 = \lambda_3 = \frac{0.9}{3} = 0.3 \text{ 人/min}$$

按 $M/M/1$ 模型重新求解本例的各问,并将结果列于表 13.4 进行比较。

表 13.4

	1个($M/M/3/\infty/\infty$)	3个($M/M/1/\infty/\infty$)
系统空闲的概率 P_0	0.074 8	0.25(每个子系统)
顾客需等待的概率	0.57	0.75
平均队长 L_s	3.95	9(整个系统)
平均排队长 L_q	1.70	2.25(每个子系统)
平均逗留时间 $W_s(\text{min})$	4.39	10
平均等待时间 $W_q(\text{min})$	1.89	7.5

由表 13.4 不难看出,在多个并列服务台的排队系统,排成单队比排成并列多队的方式具有显著的优越性。

由于计算 P_0 和各项运行指标的公式比较复杂,现已有专门的数值表可供使用。表 13.5 给出了标准的 $M/M/C$ 系统 $W_q \cdot \mu$ 数值表。

表 13.5

$W_q \cdot \mu$ $\rho = \frac{\lambda}{C\mu}$	服务台数 C			
	$C=1$	$C=2$	$C=3$	$C=4$
0.1	0.111 1	0.010 1	0.001 4	0.000 2
0.2	0.250 0	0.041 7	0.010 3	0.003 0
0.3	0.428 6	0.098 9	0.033 3	0.013 2
0.4	0.666 7	0.190 5	0.078 4	0.037 8
0.5	1.000 0	0.333 3	0.157 9	0.087 0
0.6	1.500 0	0.562 5	0.295 6	0.179 4
0.7	2.333 3	0.960 8	0.547 0	0.357 2
0.8	4.000 0	1.777 8	1.078 7	0.745 5
0.9	9.000 0	4.263 2	2.723 5	1.969 4
0.95	19.000 0	9.256 4	6.046 7	4.457 1

如果表中没有某一 ρ 值,可利用线性插值法从表内求出。设欲查的 ρ 值介于表中的 ρ_1 和 ρ_2 之间,即 $\rho_1 < \rho < \rho_2$,与 ρ 相应的 $W_q \cdot \mu$ 值可由与 ρ_1, ρ_2 相应的 $(W_q \cdot \mu)_1$ 和 $(W_q \cdot \mu)_2$ 按下面的线性插值公式算出:

$$(W_q \cdot \mu) = (W_q \cdot \mu)_1 + \frac{\rho - \rho_1}{\rho_2 - \rho_1} [(W_q \cdot \mu)_2 - (W_q \cdot \mu)_1]$$

例如在例 13.7 中, $C=3, \rho=0.75$,表中无此值。由于 $0.7 < \rho < 0.8$,由插值法,与 ρ 相应的 $W_q \cdot \mu$ 为

$$\begin{aligned} W_q \cdot \mu &= 0.547 0 + \frac{0.75 - 0.7}{0.8 - 0.7} [1.078 7 - 0.547 0] \\ &= 0.547 0 + 0.5 \times 0.531 7 = 0.812 9 \end{aligned}$$

$$W_q = \frac{0.812 9}{0.4} = 2.03 \text{ min}$$

这一结果与例中的计算有一点差异,是由于插值引起的。

例 13.8 公司在港口有一专用卸货泊位,货船按泊松流到达,平均每周到达 5 艘,卸货时间服从负指数分布,平均每周卸货 10 艘。①求平均每周的等待卸货艘数,每船的平均等待卸货时间和在港停留时间。②对于一船在港等待时间(不包括卸货时间),公司须按每周 20 000 元的标准付给损失费。因此,公司考虑除自己的专用泊位外,拟再租赁其他单位的另一泊位,那里的卸货率相同。如租赁费为每周 5 000 元,问租赁这泊位是否值得? ③公司也考虑不另租泊位,而是租用新式卸货设备代替原来的卸货设备,这样可提高平均卸货率到每周 15 艘。如租用这新式设备的租赁费也是每周 5 000 元,赔偿损失的标准不变,问租赁这新式设备是否值得? ④有人建议,可以②、③两项措施并举,但这样,在计算时将遇到困难(指对人们熟悉的公式而言),这个困难是什么呢?

解 ①此问题属 $M/M/1/\infty/\infty$ 类型, $\lambda = 5$ 艘/周, $\mu = 10$ 艘/周, $\frac{\lambda}{\mu} = \frac{5}{10} = \frac{1}{2}$, 平均每周的等待卸货艘数

$$L_q = L_s - \frac{\lambda}{\mu} = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{5}{10 - 5} - \frac{1}{2} = \frac{1}{2} = 0.5 \text{ 艘/周}$$

每船平均等待卸货时间

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{1}{2}}{5} = \frac{1}{10} \text{ 周/艘}$$

每船在港逗留时间

$$W_s = W_q + \frac{1}{\mu} = \frac{1}{10} + \frac{1}{10} = \frac{1}{5} \text{ 周/艘}$$

②若租用新泊位,相当于 $C=2, \lambda=5, \mu=10, \rho = \frac{\lambda}{2\mu} = 0.25$, 用线性插值公式查表求 $W_q \cdot \mu$, 步骤如下:

$$W_q \cdot \mu = 0.417 + \frac{0.25 - 0.2}{0.3 - 0.2} (0.0989 - 0.0417) = 0.0703$$

$$W_q = \frac{0.0703}{10} = 0.00703$$

$$L_q = W_q \cdot \lambda = 0.007 \times 5 = 0.035 \text{ 艘/周}$$

比原来减少的赔付损失费为

$$20\,000 \times (0.5 - 0.035) = 9\,300 > 5\,000$$

因此,值得租用另一泊位。

③若租用新式设备,相当于 $M/M/1/\infty/\infty, \lambda=5, \mu=15$,

$$L_q = \frac{5}{15 - 5} - \frac{5}{15} = \frac{1}{6} = 0.167 \text{ 艘/周}$$

比原来减少的赔付损失费为

$$20\,000 \times (0.5 - 0.167) = 6\,660 > 5\,000$$

因此,租用新式设备也是值得的。

④若②、③两项措施并举,相当于 $C=2$, 但两个服务台的平均服务率 μ 不相等, 与人们熟悉的公式前提不符。

13.5 一般服务时间的 M/G/1 排队模型

本节讨论具有泊松输入、一般分布的服务时间、系统容量和顾客源均无限制的单服务台排队系统。

13.5.1 M/G/1/∞/∞排队模型

假设服务时间 v 的期望 $E(v)$ 和方差 $D(v)$ 存在, 服务强度 $\rho = \lambda E(v) < 1$, 则可利用布拉切-辛钦(P-K)公式及里特公式求出系统运行指标:

$$L_s = \rho + \frac{\rho^2 + \lambda^2 D(v)}{2(1-\rho)}$$

$$W_s = \frac{L_s}{\lambda}$$

$$W_q = W_s - E(v)$$

$$L_q = \lambda W_q = L_s - \rho$$

其中 L_s 的计算公式称做 P-K 公式。它表明, 只要知道服务时间 v 的期望和方差, 而不管 v 是服从什么分布, 都可以求出系统运行指标。

例 13.9 有一汽车冲洗台, 汽车按平均每小时 18 辆的泊松流到达, 冲洗时间 v 的期望 $E(v) = 0.05\text{h/辆}$, 方差 $D(v) = 0.01\text{h/辆}$, 求系统的运行指标, 并对系统进行评价。

解 此问题属于 M/G/1/∞/∞ 类型, $\lambda = 18$, $\rho = \lambda E(v) = 0.9$, $D(v) = 0.01$, 于是有

$$L_q = \frac{18^2 \times 0.01 + (0.9)^2}{2(1-0.9)} = 20.25 \text{ 辆/h}$$

$$L_s = 20.25 + 0.9 = 21.15 \text{ 辆/h}$$

$$W_s = \frac{21.15}{18} = 1.175 \text{ h/辆}$$

$$W_q = \frac{20.15}{18} = 1.125 \text{ h/辆}$$

上述结果表明, 该服务机构存在的突出问题是顾客的平均等待时间太长, 是被服务时间的 $\frac{1.125}{0.05} = 22.5$ 倍。

13.5.2 M/D/1/∞/∞排队模型

作为 M/G/1/∞/∞ 的一种特殊情形, M/D/1/∞/∞ 表示泊松输入、定长服务时间以及系统容量和顾客源均无限制的单服务台排队系统。这里的服务时间 $v \equiv E(v)$, $D(v) = 0$, 由 P-K 公式可得

$$L_s = \rho + \frac{\rho^2}{2(1-\rho)}, L_q = L_s - \rho = \frac{\rho^2}{2(1-\rho)}$$

$$W_s = \frac{L_s}{\lambda}, W_q = \frac{L_q}{\lambda}$$

若记 $E(v) = \frac{1}{\mu}$, 则有

$$L_q = \frac{\lambda^2}{2\mu(\mu - \lambda)}, W_q = \frac{\lambda}{2(\mu - \lambda)}$$

均为标准的 M/M/1 系统相应运行指标值的一半,可见系统内部越有规律越省时间。

例 13.10 某实验室有一台自动检验机器性能的仪器,要求检验机器的顾客按泊松过程到达,平均每小时 4 个顾客,检验每台机器所需时间为 6 min。求:①在检验室内的平均机器台数 L_s ;②等待检验的平均机器台数 L_q ;③每台机器在室内的平均消耗时间 W_s ;④每台机器平均等待检验的时间 W_q 。

解 此问题属于 (M/D/1/∞/∞) 类型, $\lambda = 4$ 个/h, $E(v) = \frac{1}{10}$ h, $\rho = \frac{4}{10}$, $D(v) = 0$

$$\textcircled{1} L_s = 0.4 + \frac{(0.4)^2}{2(1-0.4)} = 0.533 \text{ 台}$$

$$\textcircled{2} L_q = 0.533 - 0.4 = 0.133 \text{ 台}$$

$$\textcircled{3} W_s = \frac{0.533}{4} = 0.133 \text{ h} = 8 \text{ min}$$

$$\textcircled{4} W_q = \frac{0.133}{4} = 0.033 \text{ h} = 2 \text{ min}$$

13.5.3 M/E_k/1/∞/∞ 排队模型

本模型也是 M/G/1/∞/∞ 的一种特殊情形,服务时间 v 服从 k 阶爱尔朗分布。其实际背景是服务机构由 k 个串联的服务台组成,顾客为接受服务须经过全部 k 个服务台。每个服务台的服务时间 v_i 均服从参数为 $k\mu$ 的负指数分布,则总共的服务时间 $v = \sum_{i=1}^k v_i$ 便服从 k 阶爱尔朗分布,且 $E(v) = \frac{1}{\mu}$, $D(v) = \frac{1}{k\mu^2}$ 。由 P-K 公式有

$$L_s = \rho + \frac{\lambda^2 \left(\frac{1}{\mu^2} + \frac{1}{k\mu^2} \right)}{2 \left(1 - \frac{\lambda}{\mu} \right)} = \rho + \frac{(k+1)\rho^2}{2k(1-\rho)}, \rho = \frac{\lambda}{\mu}$$

$$L_q = L_s - \rho = \frac{(k+1)\rho^2}{2k(1-\rho)}$$

$$W_s = \frac{L_s}{\lambda}, W_q = \frac{L_q}{\lambda}$$

例 13.11 某单人裁缝店做西服,每套需经过 4 个不同的工序,4 个工序完成后才开始做另一套。每一工序的时间服从负指数分布,平均为 2 h。顾客到来服从泊松分布,平均订货率为每周 5.5 套(设一周工作 6 天,每天 8 h)。问一顾客为等到做好一套西服平均需多长时间?

解 此问题属于 (M/E₄/1/∞/∞) 类型, $\lambda = 5.5$ 套/周, $\frac{1}{4\mu} = 2$ h, $\mu = \frac{1}{8}$ 套/h = 6 套/周, $\rho = \frac{5.5}{6}$ 。因有

$$L_s = \frac{5.5}{6} + \frac{\left(\frac{5.5}{6}\right)^2 + (5.5)^2 \times \frac{4}{6^2}}{2 \left(1 - \frac{5.5}{6} \right) \times 4} = 7.2188$$

顾客为等到做好一套西服的平均时间

$$W_s = \frac{L_s}{\lambda} = \frac{7.2188}{5.5} = 1.3 \text{ 周}$$

在 $M/E_k/1/\infty/\infty$ 中, 当 $k=1$ 时即为 $M/M/1/\infty/\infty$, 而当 $k \rightarrow \infty$ 时, 即化为 $M/D/1/\infty/\infty$ 。

本节讨论了服务时间服从一般分布的几种排队模型。对于到达和服务均为一般分布的情形, 可采用随机模拟的方法求得近似解。

13.6 排队系统的优化

排队系统的优化一般包括两方面的问题: 一是系统设计优化, 即如何设计一个排队系统可使总效益最优, 这属于静态优化问题; 二是系统控制优化, 即对于一给定的排队系统如何运营可使某目标达到最优, 这属于动态优化问题。本节只讨论前者, 即系统设计优化问题, 主要讨论 $M/M/1$ 系统的最优服务率和 $M/M/C$ 系统的最佳服务台数的确定。

13.6.1 $M/M/1$ 系统的最优服务率 μ^*

1. 标准的 $M/M/1$ 系统

设 C_s 为系统对每位顾客的单位时间服务成本; C_w 为每位顾客在系统停留单位时间的损失费用; z 为总费用。 z 由来自系统的服务成本 $C_s\mu$ 和来自顾客的等待损失 $C_w L_s$ 构成, $L_s = \frac{\lambda}{\mu - \lambda}$ 也由 μ 决定。费用曲线如图 13.9 所示。

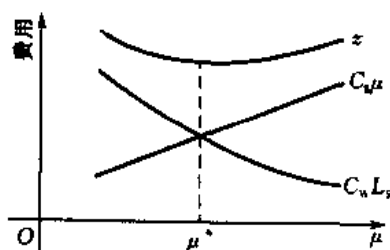


图 13.9

目标是极小化总费用 z , 即

$$\min z = C_s \mu + C_w L_s$$

令

$$\frac{dz}{d\mu} = C_s + C_w \lambda \frac{1}{(\mu - \lambda)^2} = 0$$

$$\text{解得 } \mu^* = \lambda + \sqrt{\frac{C_w}{C_s} \lambda}.$$

如果目标为系统的服务利润最大化, 仍记 z 表示总利润, 则表达式为

$$\max z = \mu(1 - P_0)G - C_s \mu$$

其中 G 为单位时间对每位顾客服务的收入, $P_0 = 1 - \frac{\lambda}{\mu}$ 。

例 13.12 设货船按泊松流到达某一港口, 平均到达率 $\lambda = 50$ 条/天, 平均卸货率为 μ 。又知船在港口停泊一天的费用为 1 货币单位, 平均卸货费为 μC_s , 其中 $C_s = 2$ 货币单位, 求使总费用最少的平均服务率 μ^* 。

解 此问题属于标准的 $M/M/1$ 系统服务率优化问题, $\lambda = 50$, $C_s = 2$, $C_w = 1$, 最佳服务率为

$$\mu^* = \lambda + \sqrt{\frac{C_w}{C_s} \lambda} = 50 + \sqrt{\frac{50}{2}} = 55 \text{ 条/天}$$

2. 系统容量有限的 M/M/1 系统

设考虑的目标是系统的服务利润最大化。用 z 表示总利润,则有

$$\max z = \lambda(1 - P_N)G - C_s\mu$$

其中 $\lambda(1 - P_N)$ 为顾客的有效到达率,即单位时间实际进入系统的顾客数(等于单位时间实际服务完的顾客数 $\mu(1 - P_0)$); $\lambda(1 - P_N)G$ 为单位时间的服务总收入。将 z 表示为

$$z = \lambda G \frac{1 - \rho^N}{1 - \rho^{N+1}} - C_s\mu = \lambda\mu G \frac{\mu^N - \lambda^N}{\mu^{N+1} - \lambda^{N+1}} - C_s\mu$$

令 $\frac{dz}{d\mu} = 0$, 得

$$\rho^{N+1} \left[\frac{N - (N+1)\rho + \rho^{N+1}}{(1 - \rho^{N+1})^2} \right] = \frac{C_s}{G}$$

当给定 N 和 $\frac{C_s}{G}$ 后,即可由上式解得 μ^* 。

实际中还有一些其他形式的 M/M/1 系统服务优化问题,可根据具体情况处理。

例 13.13 考虑一个 M/M/1/N/∞ 系统,设 $\lambda = 10$ 人/h, $\mu = 30$ 人/h, $N = 2$ 。管理者想改进服务机构,方案有两个:方案 A 是增加一个等待空间,即使 $N = 3$;方案 B 是提高平均服务率到 $\mu = 40$ 人/h。设每服务一个顾客的平均收入不变,问哪个方案将获得更大的收入? 若 λ 增加到 30 人/h,又将得到什么结果?

解 对于方案 A,单位时间内实际进入系统的顾客平均数为

$$\lambda_A = \lambda(1 - P_3) = \lambda \left(\frac{1 - \rho^3}{1 - \rho^4} \right) = 10 \left[\frac{1 - \left(\frac{1}{3}\right)^3}{1 - \left(\frac{1}{3}\right)^4} \right] = 9.75 \text{ 人}$$

对于方案 B, $\mu = 40$ 人/h,单位时间内实际进入系统的顾客平均数为

$$\lambda_B = \lambda(1 - P_2) = \lambda \left(\frac{1 - \rho^2}{1 - \rho^3} \right) = 10 \left[\frac{1 - \left(\frac{1}{4}\right)^2}{1 - \left(\frac{1}{4}\right)^3} \right] = 9.52 \text{ 人}$$

由于收入由顾客数决定,故方案 A(扩大等待空间)将获得更大的收入。

若 λ 增加到 30 人/h,有

$$\lambda_A = 30 \times \frac{3}{3+1} = 22.5 \text{ 人/h}$$

$$\lambda_B = 30 \times \frac{1 - \left(\frac{3}{4}\right)^2}{1 - \left(\frac{3}{4}\right)^3} = 22.7 \text{ 人/h}$$

故方案 B(提高服务率)将获得更大的收入。

13.6.2 M/M/C 系统的最佳服务台数 C^*

这里仅讨论标准的 M/M/C 系统。设 C_s 为每个服务台的单位时间服务成本, C_w 为每位顾客在系统停留单位时间的损失费用, z 为总成本,目标为

$$\min z = C_s \cdot C + C_w L_s$$

这里 $z = z(C)$ 是台数 C 的离散函数, 不能采用求导的方法求极值点, 可采用边际分析法。

设 C^* 是 $z(C)$ 的极小点, 则 C^* 应满足

$$\begin{cases} z(C^*) \leq z(C^* - 1) \\ z(C^*) \leq z(C^* + 1) \end{cases}$$

化简得

$$L_s(C^*) - L_s(C^* + 1) \leq \frac{C_s}{C_w} \leq L_s(C^* - 1) - L_s(C^*)$$

依次计算 $L_s(1), L_s(2), L_s(3), \dots$ 的相邻两项之差, 看常数 $\frac{C_s}{C_w}$ 落在哪二者之间, 便可确定 C^* 。

例 13.1.4 某检验中心为各工厂服务, 要求做检验的工厂(顾客)的到来服从泊松流, 平均到达率为每天 48 次, 每次来检验由于停工等原因损失为 6 元。服务(做检验)时间服从负指数分布, 平均服务率每天 25 次, 每设置 1 个检验员的服务成本(工资及设备损耗)为每天 4 元, 其他条件适合标准的 M/M/C 系统。问应设几个检验员可使总费用的平均值为最小?

解 $C_s = 4, C_w = 6, \lambda = 48, \mu = 25, \frac{\lambda}{\mu} = 1.92$, 设检验员数为 C , 根据标准的 M/M/C 模型的有关计算公式, 有

$$P_0 = \left[\sum_{k=0}^{C-1} \frac{(1.92)^k}{k!} + \frac{1}{(C-1)!} \cdot \frac{(1.92)^C}{C-1.92} \right]^{-1}$$

$$L_s = \frac{P_0 (1.92)^{C+1}}{(C-1)! (C-1.92)^2} + 1.92$$

将 $C=1, 2, 3, 4, 5$ 依次代入计算得到表 13.6。

表 13.6

检验员数 C	平均顾客数 $L_s(C)$	$L_s(C) - L_s(C+1) \sim$ $L_s(C-1) - L_s(C)$	总费用(元/天) $z(C)$
1	∞		∞
2	24.49	$21.845 \sim \infty$	154.94
3	2.645	$0.582 \sim 21.845$	27.87
4	2.063	$0.111 \sim 0.582$	28.38
5	1.952		31.71

由于 $\frac{C_s}{C_w} = \frac{4}{6} = 0.67$ 落在区间 $(0.582, 21.845)$ 内, 故 $C^* = 3$, 即当设 3 个检验员时可使总费用 z 最小。最小总费用

$$z(C^*) = z(3) = 27.87 \text{ 元。}$$

习 题

13.1 某单人理发店的顾客到达为泊松流, 平均每小时 3 人, 理发时间服从负指数分布, 平均需 10 min。求:

- (1) 理发店空闲的概率;
- (2) 店内有 4 个顾客的概率;
- (3) 店内至少有 1 个顾客的概率;
- (4) 店内顾客的平均数;

- (5)等待理发的顾客平均数;
- (6)顾客在店内平均等待时间;
- (7)顾客在店内须消耗 15 min 以上的概率。

13.2 设有一单人打字室,顾客的到达为泊松流,平均到达时间间隔为 20 min,打字时间服从负指数分布,平均为 15 min。求:

- (1)顾客来打字不必等待的概率;
- (2)打字室内顾客的平均数;
- (3)顾客在打字室内的平均逗留时间;

(4)若顾客在打字室内的平均逗留时间超过 1.25 h,则主人将考虑增加设备及打字员。问顾客的平均到达率为多少小时,主人才会考虑这样做?

13.3 汽车按平均每小时 90 辆的泊松流到达高速公路上的一个收费关卡,通过关卡的平均时间为 38 s。由于驾驶人员反映等待时间太长,主管部门打算采用新装置,使汽车通过关卡的平均时间减少到平均 30 s。但增加新装置只有在原系统中等待的汽车平均数超过 5 辆和新系统中关卡的空闲时间不超过 10% 时才是合算的。根据这一要求,分析采用新装置是否合算?

13.4 某工厂有一个半成品加工操作间,内设一个半成品加工操作台和可存放 2 个待加工半成品的场地。已知半成品按平均每天 3 个的泊松过程到达该操作间,而完成该半成品加工的必要时间服从平均每个需 $\frac{1}{4}$ 天的负指数分布。若半成品到达操作间时操作间内已没有场地存放,则要运往其他较远的地方。求:

- (1)操作间空闲的概率;
- (2)半成品到达操作间时又需运往其他地方的概率;

(3)如果每移动一个半成品到它处需 100 元,为提高效率减少移动费用,可采用两种改进方案。方案一:增加一个半成品存放场地,每天需要 10 元;方案二:提高加工效率至每个 $\frac{1}{3}$ 天,每天需 15 元。问是否应该采用改进方案,如应该则采用何种方案最佳?

13.5 在习题 13.1 中,理发店内只有 5 把椅子供人们排队等待理发。当 5 把椅子都坐满时,后来的顾客不进店就离开。再求习题 13.1 中的(1)~(6)问,并求(7)顾客来到时店内已满的概率。以本题为例比较系统容量为有限和无限两种情形的结果。

13.6 某企业有 5 台运货车辆,已知每台车每运行 100 h 平均需维修 2 次,每次需时 20 min,需维修车数和维修时间分别服从泊松和负指数分布。求该企业全部车辆正常运行的概率,及分别有 1,2,3 辆车不能正常运行的概率。

13.7 某街道口有一电话亭,在步行距离为 4 min 的拐弯处有另一电话亭。已知每次电话的平均通话时间均为 $\frac{1}{\mu} = 3$ min 的负指数分布,到达这两个电话亭的顾客均为 $\lambda = 10$ 个/h 的泊松流。假如有一名顾客去其中一个电话亭打电话,到达时正有人通话,并且还有一个人在等待,问这名顾客应在原地等待,还是转去另一电话亭打电话?

13.8 机场有两条跑道,一条专供起飞用,一条专供降落用。已知要求起飞和要求降落的飞机都分别按平均每小时 25 架次的泊松流到达,每架起飞或降落占用跑道的的时间都服从平均 2 min 的负指数分布。又设起飞和降落是彼此无关的。求:

- (1)试求一架飞机起飞或降落为等待使用跑道所需的平均时间;

(2)若机场拟调整使用跑道办法,每条跑道都可做起飞或降落用,但为了安全,每架飞机占用跑道时间延长为平均 2.16 min 的负指数分布,这时要求起飞和要求降落的飞机将混合成一个参数为 50 架次/h 的泊松到达流。试计算这情形下的平均等待时间。

- (3)以上两种方法哪个更好些呢?

13.9 一个办事员核对登记的申请书时,必须依次检查 8 张表格,核对每张表格平均需 1 min。顾客到达

率为每小时 6 人,核对时间和到达间隔均服从负指数分布。求:

(1)办事员空闲的概率;

(2) L_s, L_q, W_s, W_q 。

13.10 工件按泊松流到达某加工设备, $\lambda = 20$ 个/h,据测算该设备每多加工一个工件将增加收入 10 元,而工件每多等待或滞留一个小时将增加支出 1 元,试确定该设备的最优加工效率 μ 。

13.11 要求在机场着陆的飞机数量服从泊松分布,平均每小时 18 架次,每次着陆需占用机场跑道的平均时间为 2.5 min,服从负指数分布。试问该机场应设置多少条跑道,使要求着陆的飞机需在空中等待的概率不超过 5%,求这种情况下跑道的平均利用率。

第 14 章 马尔可夫分析

14.1 引言

马尔可夫分析法是俄国数学家马尔可夫(A. Markov)在 1907 年提出,并由蒙特-卡罗(Monte-Carlo)加以发展而建立的一种分析方法。它主要用于分析随机事件未来发展变化的趋势,即利用某一变量的现在状态和动向去预测该变量未来的状态及其动向,以预测未来某特定时期可能产生的变化,以便采取相应的对策。

为了了解马尔可夫分析用于“预测—对策—决策”的情况,首先介绍一个实例,即所谓的商标转换问题。

例 14.1 设某地居民的牛奶供应由 A、B、C 三厂负责,每月订一次,为简化问题的讨论,假定牛奶固定销售给 1 000 户顾客,要订哪厂牛奶由顾客自己选择。因广告宣传、服务质量等原因,用户会改换厂家。假设有 6 月份三个厂销售情况的市场调查记录,具体统计资料如表 14.1 所示。

表 14.1

牛奶厂	六月份顾客的变化			
	6 月 1 日顾客数	得	失	7 月 1 日顾客数
A	200	60	40	220
B	500	40	50	490
C	300	35	45	290

由表中数据可知,A 厂 6 月份从 B、C 厂得到了 60 户顾客,然而失掉了 40 户(净增 20 户);B 从 A、C 得到 40 户,然而失去 50 户(净失 10 户),C 从 A、B 得到 35 户,然而失去了 45 户(净失 10 户)。因此,A 厂 6 月份所增加的 20 户顾客实际上是这三厂原来顾客动态变化的结果,市场上称这种行为为商标转换。

牛奶厂为做好销售工作,就应当对此进行更细微的分析。下面进一步分析三厂顾客得失的详细情况,并算出它们得失的概率如表 14.2 所示。

表中 p_{ij} 表示 i 失于 j 的概率,如 B 厂 6 月份失于 A 厂 35 户,而 B 有 500 户,因此,B 厂 6 月份失于 A 的概率为

$$p_{BA} = 35/500 = 0.070$$

同理,可得

$$p_{CA} = 25/300 = 0.083, p_{AA} = 160/200 = 0.80$$

$$p_{AB} = 20/200 = 0.10, p_{BB} = 450/500 = 0.90$$

$$p_{CB} = 20/300 = 0.067, p_{AC} = 20/200 = 0.10$$

$$p_{BC} = 15/500 = 0.030, p_{CC} = 255/300 = 0.85$$

表 14.2

牛奶厂		A	B	C
6月1日顾客数		200	500	300
7月1日顾客数		220	490	290
得失 值及 其概 率	A	160	35	25
		$P_{AA}=0.80$	$P_{BA}=0.070$	$P_{CA}=0.083$
	B	20	450	20
		$P_{AB}=0.10$	$P_{BB}=0.90$	$P_{CB}=0.067$
	C	20	15	255
		$P_{AC}=0.10$	$P_{BC}=0.03$	$P_{CC}=0.85$

将上表的得失情况用概率矩阵的形式进行描述,则有

$$\begin{array}{c}
 \begin{array}{ccc}
 & A & B & C \\
 \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0.80 & 0.10 & 0.10 \\ 0.07 & 0.90 & 0.03 \\ 0.083 & 0.067 & 0.85 \end{bmatrix} & \begin{array}{c} \downarrow \\ \downarrow \\ \downarrow \end{array} & \begin{array}{c} \text{(维持和获得)} \\ \text{(维持和获得)} \\ \text{(维持和获得)} \end{array} \\
 & \xrightarrow{\hspace{1cm}} & \text{(维持和损失)}
 \end{array}
 \end{array}$$

如矩阵第一行表示 A 厂以 0.8 的概率保持原订户,以 0.1 的概率丧失原订户于 B 厂,以 0.1 的概率丧失原订户于 C 厂。第一列表示 A 厂以 0.8 的概率保持原订户,以 0.07 的概率从 B 厂获得新订户,以 0.083 的概率从 C 厂获得新订户。显然,有了这样的信息,就比较容易考察用户的转换情况。

根据以上数据可做以下工作:

- ①预测未来某个时刻各销售者在市场上占有的份额(市场占有率);
- ②预测将来销售者在市场份额的得失比率;
- ③预测市场是否会出现市场平衡状态(稳定市场份额);
- ④按对市场份额得失分析销售者的推销活动,指导生产厂家促销。

如可根据上述数据预测 8 月 1 日 A、B、C 三厂的市场占有率,则 8 月 1 日的状态为

$$[0.22 \quad 0.49 \quad 0.29] \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.07 & 0.9 & 0.03 \\ 0.083 & 0.067 & 0.85 \end{bmatrix} = [0.234 \quad 0.483 \quad 0.283]$$

如 A 厂占有份额 = $0.22 \times 0.8 + 0.07 \times 0.49 + 0.083 \times 0.29 = 0.234$

(A 厂保持率) (B 厂转入率) (C 厂转入率)

其中,向量 $[0.22 \quad 0.49 \quad 0.29]$ 为各厂 7 月份的市场占有份额(订户数与总订户数之比),则 8 月份 A 厂拥有全部顾客的 23.4%,B 厂为 48.3%,C 厂为 28.3%。

如果要预测 9 月份、10 月份甚至更远时期的市场占有率,可以像上面那样,逐月计算得到。可见,马尔可夫分析方法能为决策者提供许多有价值的参考信息。其应用范围是广阔的。

那么,马尔可夫分析主要的研究内容是什么呢?从引例知,某些事物在发展变化,现在的

状态(试验结果)只与过去前一次试验有关,而与在此以前的经过无关,则称这样的过程为马尔可夫过程。一连串(若干个)离散的马尔可夫过程的整体,称为马尔可夫链。因而,如同引例中列举的那样,对于马尔可夫链或马尔可夫过程可能产生的演变加以分析,以观察或预测其未来发展变化的趋势便是马尔可夫分析的研究内容。

由于马尔可夫过程为概率过程,所以在研究马尔可夫过程之前,首先要阐明概率向量、概率矩阵与正规概率矩阵等基本概念。

14.2 关于正规随机矩阵的一些知识

1. 概率向量

任意一个向量 $u = [u_1 \ u_2 \ \cdots \ u_n]^T$, 若 $u_i \geq 0$, 且 $\sum_{i=1}^n u_i = 1$, 则称 u 为概率向量。例

14.1 中矩阵的每一行均为一个概率向量。

2. 概率矩阵

在方阵 $P = [p_{ij}]_{n \times n}$ 中, 若各个行向量都为概率向量, 则称此方阵为概率矩阵或随机矩阵。例 14.1 中的矩阵就是一个随机矩阵。

概率矩阵具有如下性质。

性质 14.1 设 $u \in R^n$ 为一个概率向量, $A = [a_{ij}]_{n \times n}$ 是一个概率矩阵, 则 $A^T u = y$ 也是一个概率向量。

$$\begin{aligned} \text{证 } y^T &= u^T A = [u_1 \ u_2 \ \cdots \ u_n] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \\ &= \left[\sum_{i=1}^n u_i a_{i1} \quad \sum_{i=1}^n u_i a_{i2} \quad \cdots \quad \sum_{i=1}^n u_i a_{in} \right] \end{aligned}$$

则 y^T 各分量之和为

$$\sum_{j=1}^n y_j = \sum_{j=1}^n \sum_{i=1}^n u_i a_{ij} = \sum_{i=1}^n u_i \left(\sum_{j=1}^n a_{ij} \right) = \sum_{i=1}^n u_i = 1$$

所以, $A^T u$ 为一个概率向量。

性质 14.2 若 A 和 B 都为概率矩阵, 则 AB 亦为概率矩阵, A^n 亦为概率矩阵。

证 用 A_i 表示由矩阵 A 的第 i 行组成的向量, 由性质 14.1 知, $B^T A_i$ 都是概率向量, 而

$$AB = \begin{bmatrix} A_1^T B \\ \vdots \\ A_n^T B \end{bmatrix}$$

故 AB 的每一行组成的向量均为概率向量, 即 AB 为随机矩阵, 从而易证 A^n 为随机矩阵。

3. 正规概率矩阵

对于任一概率矩阵 P , 若存在 m , 使 P^m (m 为大于 1 的正整数) 的所有元素都是正数, 则称 P 为正规概率矩阵(由性质 14.2 知, P^m 必是一个随机矩阵)。

如概率矩阵 $P = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, $P^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$, 则 P 是一个正规概率矩阵。再如概率矩阵

$$Q = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, Q^2 = \begin{bmatrix} 1 & 0 \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}, Q^3 = \begin{bmatrix} 1 & 0 \\ \frac{7}{8} & \frac{1}{8} \end{bmatrix}, \dots, Q^m = \begin{bmatrix} 1 & 0 \\ \frac{2^m-1}{2^m} & \frac{1}{2^m} \end{bmatrix}, \text{则 } Q \text{ 不是正规概率}$$

矩阵。

以下给出正规随机矩阵的几个重要性质。

若 A 是一个正规随机矩阵, 则有:

- ①一定存在一个概率向量 X , 使得 $A^T X = X$ 成立, 且 X 的各分量皆为正数;
- ② A 的各次方幂 $A^1, A^2, A^3, \dots, A^k, \dots$ 组成的序列会趋近于一个固定方阵 B , 即 $A^k \rightarrow B$ (当 $k \rightarrow \infty$), 且 B 的每一行均为 X^T ;
- ③设 u 为任一 n 维概率向量, 则向量序列 $A^T u, (A^2)^T u, (A^3)^T u, \dots, (A^k)^T u, \dots$ 趋近于概率向量 X , 即有 $(A^k)^T u \rightarrow X (k \rightarrow \infty)$ 。

例 14.2 试用正规随机矩阵 $A = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ 验证上述定理。

解 设概率向量 $X = [x_1 \ x_2]^T$ 满足方程组 $A^T X = X$, 于是便得到方程组

$$\begin{cases} \frac{1}{2}x_2 = x_1 \\ x_1 + \frac{1}{2}x_2 = x_2 \end{cases}$$

不难看出, 这个方程组的两个方程不是相互独立的, 事实上, 将这两个方程相加, 可得恒等式:

$$x_1 + x_2 \equiv x_1 + x_2$$

于是, 用方程 $x_1 + x_2 = 1$ 来取代上述方程组中的第 2 个方程, 得新的方程组

$$\begin{cases} \frac{1}{2}x_2 = x_1 \\ x_1 + x_2 = 1 \end{cases}$$

解之得

$$X = [x_1 \ x_2]^T = \left[\frac{1}{3} \ \frac{2}{3} \right]^T$$

进一步, 矩阵序列 A^1, A^2, A^3, \dots 趋近于各行都以向量 X^T 所构成的方阵 B 。

$$B = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

事实上, 有

$$A^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}, A^3 = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{5}{8} \end{bmatrix}, A^4 = \begin{bmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{5}{16} & \frac{11}{16} \end{bmatrix}, A^5 = \begin{bmatrix} \frac{5}{16} & \frac{11}{16} \\ \frac{11}{32} & \frac{21}{32} \end{bmatrix},$$

$$A^6 = \begin{bmatrix} 11 & 21 \\ 32 & 32 \\ 21 & 43 \\ 64 & 64 \end{bmatrix} \cdots \cdots \rightarrow B$$

另设 $u = [u_1 \quad u_2]^T$ 为任一概率向量, 由 $A^k \rightarrow B (k \rightarrow \infty)$ 可得

$$(A^k)^T u \rightarrow B^T u$$

$$B^T u = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}(u_1 + u_2) \\ \frac{2}{3}(u_1 + u_2) \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix} = X$$

14.3 马尔可夫链

14.3.1 一般随机过程

在某些现实系统中, 表征系统特征的变量具有随机性, 而且系统状态随时间而变化, 即系统状态变量在每个时点上的取值是随机的, 那么, 对于这类系统就需用以时间为参数的随机变量来描述。系统状态的这种变化过程常称为随机过程。例如, 从时间 $t=0$ 开始记录某电话总机的呼叫次数, 设 $t=0$ 时没有呼叫, 至时刻 t 的呼叫次数记作 N_t , 则随机变量族 $\{N_t, t \geq 0\}$ 是随机过程。

下面给出随机过程的有关概念。

随机过程 (stochastic process) 是指依赖于一个变动参数 t 的一族随机变量 $\{X(t), t \in T\}$ 。变动参数 t 所有可以取值的集合 T 称为参数空间。 $X(t)$ 的值所构成的集合 S 称为随机过程的状态空间。按 S 和 T 是离散集或非离散集可将随机过程分为四类 (略)。马尔可夫分析只涉及随机过程的一个子类, 即所谓的马尔可夫过程。这类随机过程的特点是: 若已知在时间 t 系统处于状态 X 的条件下, 在时刻 $\tau (\tau > t)$ 系统所处的状态与时刻 t 以前系统所处的状态无关, 此过程便为马尔可夫过程。

例如, 在液面上放一微粒, 它由于受到大量分子的碰撞, 在液面上作不规则运动, 这就是布朗运动。由物理学知, 已知在时刻 t 的运动状态条件下, 微粒在 t 以后的运动情况和微粒在 t 以前的情况无关。若以 $X(t)$ 表示微粒在时刻 t 时的位置, 则 $X(t)$ 是马尔可夫过程。

马氏过程一般分为三类:

- ① T 连续、 S 离散的马氏过程;
- ② T 连续、 S 连续的马氏过程;
- ③ T 离散、 S 离散的马氏过程, 即马尔可夫链。

本书主要涉及马尔可夫链, 所以, 以下介绍马尔可夫链的概念。

14.3.2 马尔可夫链

定义 14.1 设 $\{X_n, n=0, 1, 2, \dots\}$ 是一个随机变量序列, 用“ $X_n = i$ ”表示时刻 n 系统处于状态 i 这一事件, 称 $p_{ij}(n) = p(X_{n+1} = j | X_n = i)$ 为在事件“ $X_n = i$ ”出现的条件下, 事件“ $X_{n+1} = j$ ”出现的条件概率, 又称它为系统的一步转移概率。若对任意的非负整数 i_1, i_2, \dots

……, i_{n-1}, i, j 及一切 $n \geq 0$, 有 $p(X_{n+1} = j | X_n = i, X_k = i_k, k = 1, 2, \dots, n-1) = p(X_{n+1} = j | X_n = i) = p_{ij}(n)$, 则称 $\{X_n\}$ 是一个马尔可夫链。

马尔可夫链的定义说明, 过程在每一时刻上的状态仅仅取决于过程在前一时刻上的状态, 而与这之前的状态无关。这一性质便是马尔可夫链的无后效性 or 无记忆性。

例如, 有一位顾客每天向一家商店买一包香烟。他购买香烟并不固定于一种牌号, 商店中 A、B、C、D、E 五种牌号的香烟他都有可能购买。设 X_m 表示他在第 m 天购买的香烟牌号。若这个人只记得昨天抽烟的味道, 以前都不记得了, 那么 X_m 取什么值, 只与 X_{m-1} 取什么值有关, 则 $\{X_m\}$ 构成一个马尔可夫链。

进一步, 若系统无论何时从状态 i 出发, 经 k 步转移到状态 j 的概率都相同, 即有下式成立:

$$p(X_{s+k} = j | X_s = i) = p(X_{k+1} = j | X_1 = i)$$

其中, i, j, k 皆为正整数, s 为任一正整数, 则称此马尔可夫链为齐次马尔可夫链。以后主要研究齐次马尔可夫链, 并略去“齐次”二字。

若系统的一步转移概率 $p_{ij}(n) = p(X_{n+1} = j | X_n = i)$ 与初始时刻 n 无关, 则可简记为 p_{ij} 。用图 14.1 的状态转移图把它表示出来。

显然, 一步转移概率具有如下性质:

$$p_{ij} \geq 0 \quad (i, j = 1, 2, \dots, n)$$

$$\sum_{j=1}^n p_{ij} = 1 \quad (i = 1, 2, \dots, n)$$

把各状态之间的一步转移概率排成矩阵

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

称 P 为转移矩阵。

每个状态 i 对应转移矩阵 P 的第 i 行。若系统处于状态 i , 则该行向量表示下一次试验的所有可能结果的概率。因此, 它是一个概率向量, 所以转移矩阵 P 是一个概率矩阵。

转移概率矩阵 P 决定了各状态间的转移规律。

例 14.3(天气预报问题) 如果明日是否有雨仅与今日天气(是否有雨)有关, 而与过去的天气无关, 并设今日下雨, 则明日有雨的概率为 α , 今日无雨则明日有雨的概率为 β ; 又假定把有雨称为 0 状态天气, 无雨称为 1 状态天气, 则本例是一个两状态的马尔可夫链。它的一步转移概率矩阵为

$$P = \begin{bmatrix} 0 & 1 \\ \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

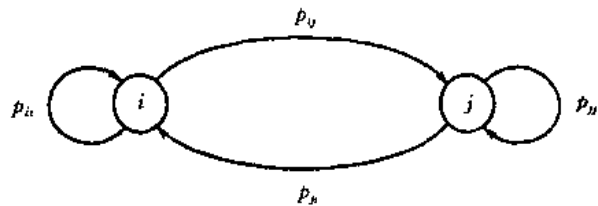


图 14.1

14.3.3 k 步转移概率与 k 步转移矩阵

p_{ij} 是系统从状态 i 一步转移到状态 j 的概率,那么,该系统从状态 i 恰好经 k 步转移到状态 j ($i \rightarrow 1 \rightarrow 2 \rightarrow \cdots \rightarrow k-1 \rightarrow j$) 的概率是多少呢?

这里,对于齐次马尔可夫链,称条件概率 $p_{ij}^{(k)} = p(X_{k+1} = j | X_1 = i)$ 为从状态 i 到状态 j 的 k 步转移概率(在我们讨论的问题里,假设状态集为有限集),并称矩阵 $P^{(k)} = (p_{ij}^{(k)})_{n \times n}$ 为 k 步转移矩阵。显然, $P^{(k)}$ 为概率矩阵,即有

$$p_{ij}^{(k)} \geq 0 \quad (i, j = 1, 2, \dots, n); \quad \sum_{j=1}^n p_{ij}^{(k)} = 1 \quad (i = 1, 2, \dots, n)$$

以下通过实例给出某些重要结论。

例 14.4 某商店对前一天来店分别购买 A、B、C 牌号香烟的顾客各 100 名的购买情况进行统计(每天都购一包),统计结果如表 14.3 所示。

表 14.3

顾客数目 前次购买品牌	今天购买情况	A	B	C
A		20	50	30
B		20	70	10
C		30	30	40

不妨认为顾客在每次购买香烟时,只对他前次所吸牌号有印象,因此,商店可采用一个齐次马氏链 $\{X_m | m = 1, 2, \dots\}$ 描述顾客对香烟的需求状况。随机变量 X_m 表示顾客在第 m 天购买的香烟牌号。为方便起见,不妨设 1、2、3 分别代表顾客购买 A、B、C 三牌号香烟的不同状态,故由表中数据不难算出,此马氏链的一步转移矩阵

$$P = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

假定一位顾客在第一天购买牌号 A 的香烟,试问他在第三天购买牌号 B 的概率是多少呢? 这实际上要求的是二步转移概率 $p_{12}^{(2)}$ 。

如设第一天购买 A 牌号,第二天购买 A、B、C 的概率分别为 $p_{11} = 0.2, p_{12} = 0.5, p_{13} = 0.3$;而第二天买 A、B、C,第三天买 B 的概率为 $p_{12} = 0.5, p_{22} = 0.7, p_{32} = 0.3$,则第三天购买 B 的概率为二步转移概率 $p_{12}^{(2)} = p(X_3 = 2 | X_1 = 1)$ 。结合 14.2 图示,根据概率乘法公式与互斥性得

$$\begin{aligned} p_{12}^{(2)} &= p_{11}p_{12} + p_{12}p_{22} + p_{13}p_{32} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \end{bmatrix} \begin{bmatrix} p_{12} \\ p_{22} \\ p_{32} \end{bmatrix} \\ &= \begin{bmatrix} 0.2 & 0.5 & 0.3 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.7 \\ 0.3 \end{bmatrix} = 0.54 \end{aligned}$$

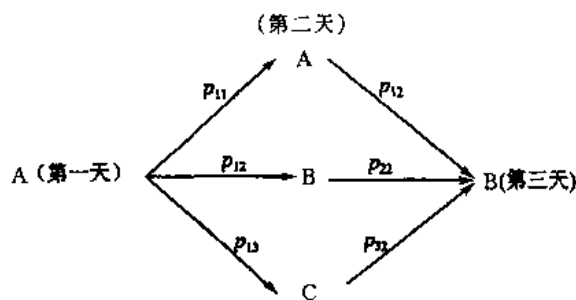


图 14.2

更一般地,结合 14.3 图示,可得齐次马氏链的二步转移概率及二步转移矩阵为

$$p_{ij}^{(2)} = \sum_k p_{ik}^{(1)} p_{kj}^{(1)}$$

$$P^{(2)} = [p_{ij}^{(2)}]_{n \times n} = P \cdot P = P^2$$

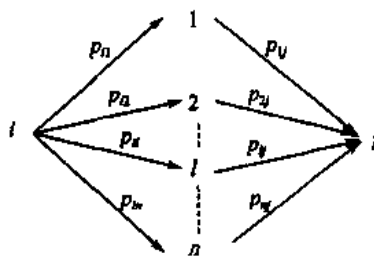


图 14.3

故知,上例中二步转移矩阵

$$P^{(2)} = P^2 = \begin{bmatrix} 0.23 & 0.54 & 0.23 \\ 0.21 & 0.62 & 0.17 \\ 0.24 & 0.48 & 0.28 \end{bmatrix}$$

进一步,还有下式成立:

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(m)} \cdot p_{kj}^{(n-m)}, \text{ 即 } P^{(n)} = P^{(m)} \cdot P^{(n-m)}$$

这个表示 n 步转移概率基本关系的方程称为切普曼-柯尔莫哥洛夫方程(Chapman-Kolmogorov equation),但这里必须在马尔可夫假设条件下方能成立。上述公式的证明有关文献已经给出,读者不妨参考。

应用切普曼-柯尔莫哥洛夫方程易知, $P^{(n)} = P^{(n-1)} \cdot P = P^n$, 即是说 n 步状态转移矩阵就等于一步状态转移矩阵的 n 次乘方。

例 14.5 如设例 14.3 中, $\alpha=0.7, \beta=0.4$, 则

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

于是,两步转移概率矩阵为

$$P^{(2)} = P \cdot P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix}$$

四步转移概率矩阵为

$$\begin{aligned} P^{(4)} &= P^{(2)} P^{(2)} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix} \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix} \\ &= \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix} \end{aligned}$$

由此可知,今日有雨第五日仍有雨的概率为 $P_{00}^{(4)} = 0.5749$ 。

如果希望得到第 k 步的状态概率,则必须首先知道过程开始时的概率分布,称之为初始概率分布,记为

$$\lambda_0 = [\lambda_1^{(0)} \quad \lambda_2^{(0)} \quad \cdots \quad \lambda_n^{(0)}]^T, \lambda_i^{(0)} = p\{X_0 = i\} \geq 0, \sum_{i=1}^n \lambda_i^{(0)} = 1$$

同样,第 k 步的概率分布表示为

$$\lambda_k = [\lambda_1^{(k)} \quad \lambda_2^{(k)} \quad \cdots \quad \lambda_n^{(k)}]^T, \lambda_j^{(k)} = p\{X_k = j\} \geq 0, \sum_{j=1}^n \lambda_j^{(k)} = 1$$

由乘法公式得

$$\lambda_j^{(k)} = p\{X_k = j\} = \sum_i p\{X_k = j | X_0 = i\} \cdot p\{X_0 = i\} = \sum_i \lambda_i^{(0)} \cdot p_{ij}^{(k)}$$

写成向量形式为: $\lambda_k = (P^k)^T \lambda_0$ 。

上式表明,第 k 步时的状态概率是初始状态概率与 k 步状态转移矩阵的乘积。一旦知道初始状态概率和一步转移矩阵,则马尔可夫链就可以完全确定了。

如在例 14.1 中,可根据转移概率矩阵 P 以及 7 月份各厂的市场占有率,通过上式,依次求得 8、9、10 月份……的市场占有率。

14.3.4 稳态概率

根据马尔可夫假设,对于一个马氏链,若已经知道初始状态概率 $\lambda_i^{(0)} = p\{X_0 = i\}$ 和一步转移矩阵 P ,就可以求出任意时刻时过程处于某种状态的概率 $\lambda_j^{(n)} = p\{X_n = j\}$ 。这时, $\lambda_j^{(n)} = p\{X_n = j\}$ 为系统瞬态概率。在实际应用当中,最感兴趣的是当 $n \rightarrow \infty$ 时,在统计平衡条件下系统所处状态的概率分布。下面讨论当 $n \rightarrow +\infty$ 时马尔可夫链的变化。

定义 14.2 若极限 $\lim_{n \rightarrow +\infty} \lambda_j^{(n)} = \lambda_j^*$ 存在,且 $\sum_j \lambda_j^* = 1$,则称 $\{\lambda_j^*\}$ 为系统的平稳分布, $\lambda^* = [\lambda_1^* \quad \lambda_2^* \quad \cdots \quad \lambda_n^*]^T$ 为系统的稳态概率向量。

稳态概率分布具有下述性质。

性质 14.3 稳态概率分布与初始概率分布无关。

性质 14.4 若马尔可夫链是标准的,即它的转移概率矩阵 P 是一个正规随机矩阵,则存在一个概率向量 $\lambda^* = [\lambda_1^* \quad \lambda_2^* \quad \cdots \quad \lambda_n^*]^T$ 满足

$$P^T \lambda^* = \lambda^*$$

λ_j^* 即为状态 j 的稳态概率, λ^* 为稳态概率向量。

显然,若转移矩阵为正规随机矩阵,则系统就必然存在平衡状态,即最终达到完全与初始状态无关的一种平衡状态。由正规随机矩阵的性质不难得出, $P^T \lambda^* = \lambda^*$ 成立。

事实上,性质 14.4 也可以如此理解:

由于

$$\lambda_k = [P^k]^T \lambda_0 = P^T \cdot [P^{k-1}]^T \lambda_0 = P^T \lambda_{k-1}$$

而随着步数的增大,有

$$\lim_{k \rightarrow +\infty} \lambda_k = \lim_{k \rightarrow +\infty} \lambda_{k-1} = \lambda^*$$

则由上式知, $\lambda^* = P^T \lambda^*$ 。

性质 14.4 同时也给出了一个求解稳态概率向量的方法,结合 $\sum_{i=1}^n x_i = 1$,有

$$\begin{cases} P^T X = X \\ \sum_{i=1}^n x_i = 1 \end{cases} \quad (1)$$

$$(2)$$

其中, $X = [x_1 \ x_2 \ \cdots \ x_n]^T$, 式(1)称为平衡方程。尽管在平衡方程中变量数与方程数相等,但在转移矩阵中必须满足每行元素之和为 1,所以应联合式(2), (用式(2)替换式(1)中某一个方程即可), 式(2)也称规范化方程。

求解上述方程组,即得稳态概率向量 X 。

例 14.6 例 14.4 中的一步转移矩阵 $P = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$, P 为正规随机矩阵,则系统

必存在惟一稳态概率向量 $X = [x_1 \ x_2 \ x_3]^T$, 由上述公式可得

$$\begin{cases} 0.2x_1 + 0.2x_2 + 0.3x_3 = x_1 \\ 0.5x_1 + 0.7x_2 + 0.3x_3 = x_2 \\ x_1 + x_2 + x_3 = 1 \end{cases}$$

解得 $X = [0.22 \ 0.57 \ 0.21]^T$, X 即为平稳状态下 A、B、C 三种牌号香烟的购买率。

综上所述,一般的齐次马尔可夫链具有如下一些性质。

①转移概率矩阵 P 是一个随机矩阵。

② n 步状态转移矩阵等于一步状态转移矩阵的 n 次乘方,即 $P^{(n)} = P^n$, 且 n 步转移概率为 $p_{ij}^{(n)} = \sum_k p_{ik}^{(m)} \cdot p_{kj}^{(n-m)}$ 。

③第 k 步的状态概率向量 λ_k 与初始概率向量 λ_0 有如下关系:

$$\lambda_k = (P^k)^T \lambda_0$$

④若转移概率矩阵 P 是一个正规随机矩阵,则系统存在一个惟一的稳态概率向量 $\lambda^* = [\lambda_1^* \ \lambda_2^* \ \cdots \ \lambda_n^*]^T$, 使得 $P^T \lambda^* = \lambda^*$, 且知:

(a) $P^{(k)} \rightarrow B$ ($k \rightarrow +\infty$) 时, B 的每一行向量相同,全为 $(\lambda^*)^T$;

(b) $\lambda_k = (P^k)^T \lambda_0 \rightarrow \lambda^*$ ($k \rightarrow +\infty$), 亦即 $\lim_{k \rightarrow +\infty} p(X_k = j) = \lambda_j^*$ 。

换言之,此时齐次马尔可夫链在经历一定时间的状态转移后,会趋于一种与初始状态无关的稳定状态。

下面是齐次马尔可夫链具体的应用实例。

14.3.5 马尔可夫分析应用实例

例 14.7 A、B、C 三家公司生产同一种产品。根据调查得知,各家公司产品上月占据市场

的份额分别为 50%、30% 和 20%。设第 m 个月顾客对这种产品的各种牌号的需求状况可用一个齐次马尔可夫链 $\{X_m | m=1, 2, \dots\}$ 表示。令状态 N_1 = 买 A 公司产品, 状态 N_2 = 买 B 公司产品, 状态 N_3 = 买 C 公司产品, 相应的概率转移矩阵

$$P = \begin{bmatrix} 0.70 & 0.10 & 0.20 \\ 0.10 & 0.80 & 0.10 \\ 0.05 & 0.05 & 0.90 \end{bmatrix}$$

试问: 本月和下月三家公司产品占据市场的份额各为多少? 一年后三家公司产品占据市场的份额各为多少?

解 上月购买 A 公司产品的顾客中, 本月将有 70% 仍买 A 公司产品, 上月购买 B 公司产品的顾客中, 本月将有 10% 转买 A 公司产品; 上月购买 C 公司产品的顾客中, 本月将有 5% 转买 A 公司产品。因而 A 公司产品本月占据市场的份额为

$$\begin{aligned} & 0.5 \times 0.70 + 0.30 \times 0.10 + 0.2 \times 0.05 \\ &= [0.5 \quad 0.3 \quad 0.2] \begin{bmatrix} 0.70 \\ 0.10 \\ 0.05 \end{bmatrix} = 0.39 \end{aligned}$$

其中, 初始概率向量 $[0.5 \quad 0.3 \quad 0.2]^T$ 表示上月三家公司产品占据市场的份额, 于是描述本月各公司占据市场份额的概率向量为

$$[0.5 \quad 0.3 \quad 0.2]P = [0.39 \quad 0.30 \quad 0.31]$$

描述下月各公司产品占据市场份额的概率向量为

$$[0.39 \quad 0.30 \quad 0.31]P = (0.5 \quad 0.3 \quad 0.2)P^2 = (0.32 \quad 0.29 \quad 0.39)$$

而描述一年后各公司产品占据市场份额的概率向量为

$$[0.5 \quad 0.3 \quad 0.2]P^{12}$$

因为 P 是一个正规随机矩阵, 故知此概率向量近似于稳态概率向量 λ^* , 求解方程

$$\begin{cases} 0.7\lambda_1 + 0.1\lambda_2 + 0.05\lambda_3 = \lambda_1 \\ 0.1\lambda_1 + 0.8\lambda_2 + 0.05\lambda_3 = \lambda_2 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1 \end{cases}$$

可得 $\lambda^* = [0.18 \quad 0.23 \quad 0.59]^T$ 。故 A 公司产品占据市场的份额一年后将由年初的 50% 下降到 18%, 情况堪忧。A 公司必须考虑改进措施。

例 14.8 (订货决策) 某商店经营一种易腐食品, 出售后一个单位可获利 $a = 5$ 元。若当天售不出去, 则每单位损失 $b = 3$ 元。该店经理统计了连续 40 天的需求情况 (不是实际销售量)。现将所得数据列出如下:

3, 3, 4, 2, 2, 4, 2, 3, 4, 4, 4, 3, 2, 4, 2, 3, 3, 4, 2, 2, 4, 3, 4, 3, 2, 3, 4, 2, 3, 2, 2, 3, 4, 2, 4, 4, 3, 2, 3, 3

经理想应用马尔可夫链来预测需求量, 确定明天进货量。① 已知当天需求量为 3 个单位, 明日应进货多少个单位? ② 若不知当天需求量, 明日应进货多少个单位?

解 令状态 N_1 = 需求量为 2 个单位, 状态 N_2 = 需求量为 3 个单位, 状态 N_3 = 需求量为 4 个单位。状态转移的情况是:

$$N_1 \rightarrow N_1, 3 \text{ 次}; N_1 \rightarrow N_2, 6 \text{ 次}; N_1 \rightarrow N_3, 4 \text{ 次}$$

$N_2 \rightarrow N_1$, 4 次; $N_2 \rightarrow N_2$, 3 次; $N_2 \rightarrow N_3$, 6 次

$N_3 \rightarrow N_1$, 6 次; $N_3 \rightarrow N_2$, 4 次; $N_3 \rightarrow N_3$, 3 次

于是, 此马尔可夫链的转移概率矩阵

$$P = \begin{bmatrix} \frac{3}{13} & \frac{6}{13} & \frac{4}{13} \\ \frac{4}{13} & \frac{3}{13} & \frac{6}{13} \\ \frac{6}{13} & \frac{4}{13} & \frac{3}{13} \end{bmatrix}$$

求解方程组

$$\begin{cases} \frac{3}{13}x_1 + \frac{4}{13}x_2 + \frac{6}{13}x_3 = x_1 \\ \frac{6}{13}x_1 + \frac{3}{13}x_2 + \frac{4}{13}x_3 = x_2 \\ x_1 + x_2 + x_3 = 1 \end{cases}$$

得稳态概率向量

$$X = \left[\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right]^T$$

这样, 应用边际分析的思想就可回答经理的问题。由于

$$\frac{b}{a+b} = \frac{3}{5+3} = \frac{3}{8} = 0.375$$

而 $p(\text{明日需求量}=4 | \text{当天需求量}=3) = p_{23} = \frac{6}{13} \approx 0.462 > \frac{b}{a+b}$, 故对①的回答为: 明日应进货 4 个单位。又因为

$$p(\text{明日需求量}=3) = x_2 = \frac{1}{3} < \frac{b}{a+b}$$

$$p(\text{明日需求量} \geq 3) = x_2 + x_3 = \frac{2}{3} > \frac{b}{a+b}$$

所以明日应进货 3 个单位。

14.4 吸收马尔可夫链

14.4.1 吸收马尔可夫链

以下研讨齐次马尔可夫链中的一种特殊类型——齐次吸收马尔可夫链以及它在经济管理问题中的具体应用。为此, 首先介绍有关的基本概念。

定义 14.3 对于马尔可夫链的状态 i , 如果 $p_{ii} = 1$, 即到达状态 i 后, 永久停留在 i , 不可能再转移到其他任何状态, 那么, 就称状态 i 为吸收状态或称为吸收态, 否则为非吸收态。

定义 14.4 若一个马氏链至少有一个吸收态, 且任何一个非吸收态到吸收态是可能的 (不必是一步), 则称此马氏链为吸收马尔可夫链。

例 14.9 甲、乙两人进行比赛, 每局比赛中甲胜的概率是 p , 乙胜的概率是 q , 和局概率是

$r(p+q+r=1)$ 。每局赛后,胜者记“+1”分,负者记“-1”分,和局不记分,当有一人获得2分时结束比赛。以 X_n 表示比赛到第 n 局时甲所得的分数,则 $\{X_n, n=1,2,\dots\}$ 就是一个吸收马氏链。

事实上,它共有五个状态,状态空间 $I=\{-2,-1,0,1,2\}$,一步转移概率矩阵

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & r & p & 0 & 0 \\ 0 & q & r & p & 0 \\ 0 & 0 & q & r & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

其中, $p_{11}=1, p_{55}=1$,这表明状态 1、5 都是吸收态。这里状态 1 意味着甲得 -2 分,甲输,比赛结束。因此,可认为 X_n 一直停留在状态 1,状态 5 也有类似的解释。由题意知,其余三个非吸收态都可能经若干步转移后到达吸收态。

当一过程到达吸收态,称它“被吸收”。可以证明:吸收马氏链将被吸收的概率为 1,或说吸收马氏链 n 步后,到达非吸收态的概率趋向于零,具体证法从略。

对于吸收马氏链,感兴趣的是如下三个问题:

①过程被吸收前,在非吸收态之间转移的平均次数是多少?

②过程从非吸收态出发到达吸收态的平均步数是多少?

③过程从非吸收态出发最终进入吸收态的概率是多少?

为此,首先考察吸收马氏链的 n 步转移概率矩阵。

事实上,对于一个有 r 个吸收态和 s 个非吸收态的吸收马氏链,经过适当排列(将吸收态集中在一起排列在前面)的一步转移概率矩阵 P 总可以表示为如下的标准形式:

$$P = \begin{matrix} & \begin{matrix} r \text{ 个吸收态} & s \text{ 个非吸收态} \end{matrix} \\ \begin{matrix} r \text{ 个吸收态} \\ s \text{ 个非吸收态} \end{matrix} & \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix} \end{matrix}$$

这是一个分块矩阵。其中,子阵 I 是一个 $r \times r$ 阶单位矩阵,它的元素是吸收态之间的转移概率;子阵 0 是一个 $r \times s$ 阶零矩阵,它的元素是吸收态到非吸收态的转移概率;子阵 R 是一个 $s \times r$ 阶子阵,它的元素是非吸收态到吸收态的转移概率;子阵 Q 是一个 $s \times s$ 阶子阵,它的元素是非吸收态之间的转移概率。

如此,可以容易地得到 n 步转移概率矩阵 $P^{(n)}$ 的分块形式如下:

$$\begin{aligned} P^{(2)} &= \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix} \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix} = \begin{bmatrix} I & 0 \\ QR + R & Q^2 \end{bmatrix} \\ P^{(3)} &= P^3 = \begin{bmatrix} I & 0 \\ Q^2R + QR + R & Q^3 \end{bmatrix} \\ &\vdots \\ P^{(n)} &= P^n = \begin{bmatrix} I & 0 \\ [Q^{n-1} + Q^{n-2} + \dots + I]R & Q^n \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ [I - Q]^{-1} \cdot [I - Q^n]R & Q^n \end{bmatrix} \end{aligned}$$

式中, Q^n 表示非吸收态之间 n 步转移概率矩阵, $[I - Q]^{-1} \cdot [I^n - Q^n]R$ 表示由非吸收态经 n 步到吸收态的转移概率矩阵。另外, 由于 n 步后过程到达非吸收态的概率趋向于零, 于是当 n 趋向于无穷时, Q^n 的每一个元素必趋向于零, 即

$$\lim_{n \rightarrow \infty} P^n = \begin{bmatrix} I & 0 \\ [I - Q]^{-1}R & 0 \end{bmatrix}$$

上式表示过程全被吸收, 而 $[I - Q]^{-1}R$ 的元素表示过程目前处于非吸收态, 最终进入吸收态的转移概率。本书将在后面详细讨论。

记矩阵 $N = [I - Q]^{-1}$, 并称它为吸收马氏链的基本矩阵, 又称特征量。显然, 基本矩阵 N 的元素给出了过程被吸收前从一个非吸收态出发, 转移到每一个非吸收态的平均次(步)数。

例 14.10 一物体作线性运动, 每次它以概率 $\frac{1}{2}$ 向右移动一单位, 或以概率 $\frac{1}{2}$ 向左移动。设置障碍后, 若物体任何时候到达这些障碍之一它将留在那儿。令状态为 0、1、2、3、4。状态 0、4 是吸收态, 其余为非吸收态, 且从中任一个到达吸收态是可能的。因此这是吸收马氏链, 它的转移概率矩阵

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

P 的标准形式为

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

$$\text{其中, } Q = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \quad \text{于是, } I - Q = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{bmatrix}$$

$$N = [I - Q]^{-1} = \begin{bmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ 1 & 2 & 1 \\ \frac{1}{2} & 1 & \frac{3}{2} \end{bmatrix}$$

由 N 知,从状态 2 出发,在吸收之前到达状态 1 的平均步数为 1,到达状态 2 的平均步数是 2,到达状态 3 的平均步数是 1。

若将 N 中某一行所有元素相加,就可得从某一非吸收态出发,在被吸收前到达各个非吸收态的平均次(步)数之和。这个值就是从该非吸收态出发到吸收时步数的平均数。这一结论可以具体描述如下:对于一个具有非吸收态的吸收马氏链,令 c 是有 s (非吸收态个数) 个分量为 1 的列向量,则向量 $t = Nc$ 具有的各个分量分别是从各个相应的非吸收态出发到被吸收时的平均步数。

例 14.11 续例 14.10,求 t 。

解

$$t = Nc = \begin{bmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ 1 & 2 & 1 \\ \frac{1}{2} & 1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$$

于是,从状态 1 开始到吸收的平均步数是 3,从状态 2 开始到吸收的平均步数是 4,从状态 3 开始到吸收的平均步数也是 3。由于此过程需从状态 2 转移至状态 1 或 3,才可到达吸收态 0 或 4,因而,从状态 2 开始比从状态 1 或 3 开始需要多一步是显然的。

最后讨论什么是一个吸收链从某一个非吸收态开始最终进入吸收态的概率。由前面推得的 n 步转移概率矩阵 P^n 不难得出:若令 b_{ij} 是一个吸收马氏链开始在非吸收态 i ,将被吸收在状态 j 的概率。令 B 是具有元素 b_{ij} 的矩阵,则 $B = NR$ 。

这可由前述 n 步转移矩阵 P^n 的极限形式得到。

事实上,从非吸收态 i 转移到吸收态 j ,可以是一步转移,转移概率是 p_{ij} ;也可以通过中间状态,先从 i 到 k (非吸收态),再到 j ,故可得方程

$$b_{ij} = p_{ij} + \sum_k p_{ik} b_{kj}$$

写成矩阵形式为

$$B = R + QB$$

$$[I - Q]B = R$$

$$B = [I - Q]^{-1}R = NR$$

例 14.12 续例 14.10、例 14.11,由 N 及 R ,立即得

$$B = NR = \begin{bmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ 1 & 2 & 1 \\ \frac{1}{2} & 1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

于是,如从状态 1 出发,可知在状态 0 吸收的概率为 $\frac{3}{4}$, 在状态 4 吸收的概率为 $\frac{1}{4}$ 。

综上所述,矩阵 N 本身回答了问题①,它给出了依赖于开始状态的过程被吸收前到达每个非吸收态的平均次数;列向量 $t = Nc$ 回答了问题②,它给出了依赖于开始状态的吸收前的平均步数;矩阵 $B = NR$ 回答了问题③,它给出了依赖于开始状态的在每个吸收态被吸收的概率。

例 14.13(企业经营状况分析) 某地企业管理部门为掌握企业经营状况的变化规律,对有关企业作了一次跟踪调查统计,得如下统计表 14.4:

表 14.4

		去年经营状况					Σ (求和)
		好	中	差	兼并	破产倒闭	
前 年 经 营 状 况	好	60	30	10	0	0	100
	中	20	60	18	2	0	100
	差	10	30	50	7	3	100
Σ (求和)		90	120	78	9	3	300

表中数字说明:100 家前年经营状况好的企业,到去年仍有 60 家保持好的状况;有 30 家变为中等;有 10 家转为差的,余类推。且马氏链(经营条件稳定,不妨假定为齐次马氏链)有五个状态,其中,“兼并”与“破产倒闭”为两个吸收态,其余为非吸收态。

由上表可得经营状况的转移矩阵如下:

$$P = \begin{matrix} & \begin{matrix} \text{兼并} & \text{破产} & \text{好} & \text{中} & \text{差} \end{matrix} \\ \begin{matrix} \text{兼并} \\ \text{破产} \\ \text{好} \\ \text{中} \\ \text{差} \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0.3 & 0.1 \\ 0.02 & 0 & 0.2 & 0.6 & 0.18 \\ 0.07 & 0.03 & 0.1 & 0.3 & 0.5 \end{bmatrix} \end{matrix} = \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix}$$

由

$$P^{(n)} = P^n = \begin{bmatrix} I & 0 \\ [Q^{n-1} + Q^{n-2} + \dots + I]R & Q^n \end{bmatrix}$$

易求 n 步转移概率矩阵 $P^{(n)}$ 。如可求得

$$P^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0.032\ 28 & 0.007\ 92 & 0.352\ 4 & 0.412\ 2 & 0.195\ 2 \\ 0.069\ 34 & 0.011\ 94 & 0.271\ 4 & 0.427\ 2 & 0.220\ 12 \\ 0.140\ 18 & 0.054\ 42 & 0.205\ 4 & 0.361\ 2 & 0.238\ 8 \end{bmatrix}$$

由 $P^{(3)}$ 可见,现在经营状况“好”的企业,经过三年运营后仍保持“好”的概率为 35.24%;变为“中等”的概率为 41.22%……而转为“破产倒闭”的概率大约只有 0.792%,余类推。

要想知道企业最终进入“兼并”或“倒闭”的概率,需计算矩阵 $B = [I - Q]^{-1}R$ 。故先求 $[I - Q]^{-1}$,得

$$[I - Q]^{-1} = \begin{bmatrix} 11.23 & 13.85 & 7.23 \\ 9.08 & 14.62 & 7.08 \\ 7.69 & 11.54 & 7.69 \end{bmatrix}$$

则

$$B = [I - Q]^{-1}R = \begin{bmatrix} 0.7831 & 0.2169 \\ 0.788 & 0.2124 \\ 0.7691 & 0.2307 \end{bmatrix}$$

由矩阵 B 知,若企业现在处于“好”的状况,则最终进入“兼并”的可能性为 78.31%;最终进入“倒闭”的可能性为 21.69%,余类推。

另外,由基本矩阵 $N = [I - Q]^{-1}$ 还可以知道以下事实。若企业现在处在“好”的状态,那么在最终进入“吸收”状态之前,将可能在“好”状态保持 11.23 年(步);在“中”状态度过 13.85 年;而在“差”的状态维持 7.23 年。(当然,这些数字都是平均值,并且不一定是连续地“保持”或“维持”)余类推。

最后,由 $t = NC$ 得($s=3$):

$$t = \begin{bmatrix} 11.23 & 13.85 & 7.23 \\ 9.08 & 14.62 & 7.08 \\ 7.69 & 11.54 & 7.69 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 32.31 \\ 30.78 \\ 26.92 \end{bmatrix}$$

由 t 知,企业由“好”状态出发到“兼并”或“倒闭”为止,平均将经过 32.31 年(步);由“中”出发只需 30.78 年;由“差”出发只需 26.92 年。

14.4.2 银行不良债务分析的马尔可夫模型

本模型使用 Markov 分析方法建立一个描述我国银行债务状态变化的模型,用此模型对银行不良债务变化趋势进行分析和预测。

1. 不良贷款状态划分

为了研究分析银行不良贷款变化情况,首先要对银行不良贷款状态进行有效划分。目前对银行不良贷款状态划分,在国外和国内有不同的口径。

为了便于分析,按以下方式对贷款状态进行划分:

N_1 ——逾期贷款,拖延支付本金利息达 0~60 天;

N_2 ——怀疑贷款,拖延支付本金利息达 61~180 天;

N_3 ——呆滞贷款,拖延支付本金利息达 181~360 天;

N_4 ——呆账贷款,拖延支付本金利息超过 360 天以上;

N_5 ——付清本金和利息。

2. 不良贷款分析的 Markov 模型

(1) 模型结构

1) X_t 随机变量,表示第 t 月时贷款。

2) t 时间(月), $t=0,1,2,\dots$ 。

3) S 状态集合, $S = \{N_1, N_2, N_3, N_4, N_5\}$, 其中 N_1, N_2, N_3, N_4, N_5 为前面所规定的贷款状态。

根据上述状态划分,上述定义的逾期贷款、怀疑贷款和呆滞贷款显然属于非吸收状态,呆账与付清是两种吸收状态,则对于状态集合 $S = \{N_1, N_2, N_3, N_4, N_5\}$ 可划分为两大类:

$$T = \{N_1, N_2, N_3\} \text{ (非吸收状态)}$$

$$A = \{N_4, N_5\} \text{ (吸收状态)}$$

4) 状态转移矩阵 为了动态研究贷款状态随时间 t 变化情况,用转移矩阵 P 来表示每隔 1 月的各种状态转移情况,有

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix}$$

其中 p_{ij} 表示当前处于 N_i 状态的贷款,1 月后将处于 N_j 的概率,并有

$$p_{ij} \geq 0, \sum_{j=1}^5 p_{ij} = 1 \quad (i=1,2,3,4,5; j=1,2,3,4,5)$$

根据 N_4, N_5 是吸收状态,可得

$$P = \begin{matrix} & \begin{matrix} N_1 & N_2 & N_3 & N_4 & N_5 \end{matrix} \\ \begin{matrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \end{matrix} & \begin{bmatrix} p_{11} & p_{12} & 0 & 0 & p_{15} \\ p_{21} & p_{22} & p_{23} & 0 & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} = \begin{matrix} & T & A \\ \begin{matrix} T \\ A \end{matrix} & \begin{bmatrix} Q & R \\ O & I \end{bmatrix} \end{matrix}$$

其中

$$Q = \begin{bmatrix} p_{11} & p_{12} & 0 \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}; \quad R = \begin{bmatrix} 0 & p_{15} \\ 0 & p_{25} \\ p_{34} & p_{35} \end{bmatrix}; \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

在上述转移矩阵中,当 $i > j, i=2,3; j=1,2$ 时, p_{ij} 应理解为当前处 N_i 状态,而在 1 月中借款者付还了部分的本金和利息,使贷款转为 N_j 状态,例如对 p_{32} ,当前贷款是处于已欠 61~180 天状态,而在本月中贷款者付还了部分贷款本金和利息,认为这贷款转为 0~60 天。 p_{32} 表示这种转移的概率。

(2) 计算步骤

1) 计算特征量(基本矩阵)

$$N = [I - Q]^{-1} = \begin{bmatrix} 1 - p_{11} & -p_{12} & 0 \\ -p_{21} & 1 - p_{22} & -p_{23} \\ p_{31} & -p_{32} & 1 - p_{33} \end{bmatrix}^{-1} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

2) 计算矩阵

$$N \cdot R = \begin{bmatrix} 1 - p_{11} & -p_{12} & 0 \\ -p_{21} & 1 - p_{22} & -p_{23} \\ -p_{31} & -p_{32} & 1 - p_{33} \end{bmatrix}^{-1} \begin{bmatrix} 0 & p_{15} \\ 0 & p_{25} \\ p_{34} & p_{35} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

3) 计算

$$[D \ G] = [\lambda_1^0 \ \lambda_2^0 \ \lambda_3^0] \mathbf{N} \mathbf{R}$$

D 为基期贷款中最终转为呆账的数目; G 为最终付清的贷款数目。

(3) 分析

k 期各种状态贷款数, 可根据 k 步转移矩阵得到, 即

$$\lambda_k = \lambda_0 \cdot \mathbf{P}^k$$

$\lambda_k = [\lambda_1^k \ \lambda_2^k \ \lambda_3^k \ \lambda_4^k \ \lambda_5^k]$ 为 k 时期各种状态贷款数 ($k = 1, 2, \dots$), 其中 $\lambda_0 = [\lambda_1^0 \ \lambda_2^0 \ \lambda_3^0 \ 0 \ 0]$ 为期初各种状态贷款数。

当 $k \rightarrow \infty$ 的极限情况时有

$$\begin{aligned} \lim_{k \rightarrow \infty} \lambda_k &= \lim_{k \rightarrow \infty} \lambda_0 \cdot \mathbf{P}^k = \lambda_0 \begin{bmatrix} \mathbf{0} & (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &= [\lambda_1^0 \ \lambda_2^0 \ \lambda_3^0 \ 0 \ 0] \cdot \begin{bmatrix} 0 & 0 & 0 & b_{11} & b_{12} \\ 0 & 0 & 0 & b_{21} & b_{22} \\ 0 & 0 & 0 & b_{31} & b_{32} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = [0 \ 0 \ 0 \ \lambda_4^* \ \lambda_5^*] \end{aligned}$$

显然有

$$[\lambda_4^* \ \lambda_5^*] = [D \ G] = [\lambda_1^0 \ \lambda_2^0 \ \lambda_3^0] \mathbf{N} \mathbf{R} \text{ 并且可得}$$

1) 各种状态贷款转为呆账及付清的概率 通过模型计算, 根据 Markov 链特征量性质可得:

b_{11} ——逾期贷款转为呆账的概率; b_{12} ——逾期贷款付清的概率

b_{21} ——怀疑贷款转为呆账的概率; b_{22} ——怀疑贷款付清的概率

b_{31} ——呆滞贷款转为呆账的概率; b_{32} ——呆滞贷款付清的概率

2) 各种状态贷款转为呆账及付清的时间 根据 Markov 链特征量 $\mathbf{N} = (m_{il})$, $i \in T$, $l \in T$, 矩阵中每一个元素 m_{il} 表示由非吸收状态 N_i 出发直至被吸收, 其中经过非吸收状态 l 的平均次数(月), 从而推出 $T(N_i) = \sum_{l \in T} m_{il}$, 表示从 N_i 状态出发至被吸收, 其中经过所有非吸收状态的次数(月)总和, 也即为状态 N_i 至被吸收所需时间, 则有:

$T(N_1) = m_{11} + m_{12} + m_{13}$ 表示逾期贷款转化呆账或付清的平均时间;

$T(N_2) = m_{21} + m_{22} + m_{23}$ 表示怀疑贷款转化呆账或付清的平均时间;

$T(N_3) = m_{31} + m_{32} + m_{33}$ 表示呆滞贷款转化呆账或付清的平均时间。

例 14.14 某银行当前借出贷款总额为 470 万, 其中属于 N_1 状态 200 万, 属于 N_2 状态 150 万, N_3 状态 120 万。而根据隔月账面变化情况分析, 近似得状态转移矩阵

$$\mathbf{P} = \begin{matrix} & \begin{matrix} N_1 & N_2 & N_3 & N_4 & N_5 \end{matrix} \\ \begin{matrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \end{matrix} & \begin{bmatrix} 0.3 & 0.3 & 0 & 0 & 0.4 \\ 0.15 & 0.25 & 0.3 & 0 & 0.3 \\ 0.1 & 0.1 & 0.3 & 0.35 & 0.15 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

其中, N_4 是呆账状态, N_5 是付清状态。根据上述信息, 对贷款状态变化和趋势进行分析。

(1) 计算特征量

$$Q = \begin{bmatrix} 0.3 & 0.3 & 0 \\ 0.15 & 0.25 & 0.3 \\ 0.1 & 0.1 & 0.3 \end{bmatrix} \quad R = \begin{bmatrix} 0 & 0.4 \\ 0 & 0.3 \\ 0.35 & 0.15 \end{bmatrix}$$

$$N = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = [I - Q]^{-1} = \begin{bmatrix} 0.7 & -0.3 & 0 \\ -0.15 & 0.75 & -0.3 \\ -0.1 & -0.1 & 0.7 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1.618 & 0.686 & 0.294 \\ 0.441 & 1.601 & 0.686 \\ 0.294 & 0.327 & 1.569 \end{bmatrix}$$

(2) 计算矩阵

$$N \cdot R = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} 1.618 & 0.686 & 0.294 \\ 0.441 & 1.601 & 0.686 \\ 0.294 & 0.327 & 1.569 \end{bmatrix} \begin{bmatrix} 0 & 0.4 \\ 0 & 0.3 \\ 0.35 & 0.15 \end{bmatrix} = \begin{bmatrix} 0.1029 & 0.8971 \\ 0.240 & 0.760 \\ 0.549 & 0.451 \end{bmatrix}$$

(3) 计算

$$[D, G] = [\lambda_1^0 \quad \lambda_2^0 \quad \lambda_3^0] NR$$

$$= [200 \quad 150 \quad 120] \begin{bmatrix} 0.1029 & 0.8971 \\ 0.240 & 0.760 \\ 0.549 & 0.451 \end{bmatrix}$$

$$= [122.46 \quad 247.54]$$

根据计算结果可知, 按目前这样的变化, 现有贷款 470 万中, 最终可能有 122.46 万将成为呆账, 收不回来, 只有 347.54 万可能收回。

逾期贷款 200 万中, 有 10.29% 可能转为呆账; 89.71% 可能付清。

怀疑贷款 150 万中, 有 24% 可能转为呆账; 76% 可能付清。

呆滞贷款 120 万中, 有 54.9% 可能转为呆账; 45.1% 可能付清。

习 题

14.1 一所大学的升级情况预测可应用一种马尔可夫模型进行。其中有 85% 的一年级学生升到二年级, 90% 二年级学生升到三年级, 92% 三年级学生升到四年级, 95% 四年级学生毕业。还知道每个年级中退学 (及转入其他大学) 的百分比为一年级 10%, 二年级 5%, 三年级 3%, 四年级 1%, 其余学生在下学年仍留在原来年级学习。请问一位刚入学的一年级学生在四年内退学的概率是多少?

14.2 老 K 公司生产的某家电产品与其他两家生产同类产品的 A 公司和 B 公司相竞争。计划用加强广告宣传以增加该产品的市场占有率。今拟订了两个广告宣传方案 D_1 和 D_2 , 公司决策者在两个条件相同 (指初始市场占有率和初始转移概率矩阵均相同) 的地区试用这两个广告方案, 试验结果得到不同的转移概率矩阵。今已知市场占有率的全国水平是: 老 K 公司产品为 28%, A 公司产品为 39%, B 公司产品为 33%。

现在要问: (1) 用初始转移概率矩阵确定在稳定 (平衡) 时, 该两试验地区市场占有率是否接近全国水平? (2) 假定两个广告宣传方案 D_1 和 D_2 成本相同, 在稳定 (平衡) 时哪种广告宣传能有最高的市场占有率?

已知初始转移概率矩阵为

$$\begin{array}{c} K \quad A \quad B \\ \begin{array}{l} K \\ A \\ B \end{array} \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \end{array}$$

广告 D_1 推出后的转移概率矩阵为

$$\begin{array}{c} K \quad A \quad B \\ \begin{array}{l} K \\ A \\ B \end{array} \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \end{array}$$

广告 D_2 推出后的转移概率矩阵

$$\begin{array}{c} K \quad A \quad B \\ \begin{array}{l} K \\ A \\ B \end{array} \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix} \end{array}$$

14.3 某供应特需商品的商店,每周在周末营业一天。该店对某种不经常有人购买的商品库存采用下述订货策略:“如结存 0 件或 1 件时,则一次订购 3 件,如结存超过 1 件时就不订购。”凡在周末停止营业时订购的商品是为了准备在下周末出售的。这一订货策略保证商品的初始库存量不是 2 件、3 件就是 4 件。

又根据统计,该商品每周的需求量为 0、1、2 和 3 件的概率分别为 0.4、0.3、0.2 和 0.1,试建立一个转移概率矩阵,用以说明由本周初始库存状态转为下周初始库存状态的概率。在达到稳定(平衡)条件下,确定库存量为 2、3 和 4 的概率。

第 15 章 随机模拟技术

15.1 模拟过程的主要步骤

模拟又称仿真,是一种基于数值方法对系统进行分析的技术。它首先为所要研究的系统设计一个模型,通过实验对系统状态的变化进行观察和统计,从而得到系统的基本性能。模拟是在管理中应用最广泛的计量方法之一。当用定量方法研究经济管理问题时,遇到以下某种情况:

①现实系统比较复杂或存在许多随机因素,各部分或因素之间的关系不能用数学模型去表达;

②有些问题虽然能够建立数学模型,但由于模型中随机因素很多,难于用解析方法求解;

③实际的环境难以观察到;

④观察一个系统可能代价昂贵;

⑤不可能有足够的时间来观察系统;

⑥对系统的实际运行或观察造成太大的破坏性,

建立和运用有效的模拟模型需要很多技巧,为了使模拟得出有意义的结论,需要作长时间的复杂分析。尽管如此,模拟仍是一种有效的方法,尤其适用于复杂的运筹学研究和系统分析。

有效的模拟需要大量的计划和组织工作,模拟过程一般必须通过下列步骤实现。

1. 问题识别

首先要对系统本身进行研究和分析,明确求解问题的目标,并决定是否要用模拟方法解决问题。

2. 建立模型

根据问题的目标和所研究的对象,建立适合于模拟求解的模型。被确定的模型应成为对真实系统的合理抽象,并将所要用的模型化为公式。有两类常用的模拟模型,即事件间隔模型和固定时间增量模型。前一种模型是指其状态仅在事件发生的时刻变化。例如研究一部机器加工作业的延迟,可用点随机过程描述:令 S_n 表示作业 n 的加工时间; A_n 表示作业 n 与 $n+1$ 的到达间隔; W_n 表示作业 n 的等待时间,那么,当 $n=0,1,2,\dots$ 时,便有

$$W_{n+1} = \max[W_n + S_n - A_n, 0] \quad (15.1)$$

一旦给定 W_{n+1} 、 S_n 与 A_n , 等待时间的序列即可产生出来。后一种模型是指其状态在任意相等的时间间隔发生变化。例如在排序问题中,一天可分成若干 15 min 的间隔,用 $A(t)$ 表示 t 到 $t+15$ 新需求数, $t=0,15,30,\dots$, 同时,完成加工的作业数为 $C(t)$, 则从 t 到 $t+15$ 净增加的作业数为 $Z(t) = A(t) - C(t)$ 。如果 t 时的作业数为 $Q(t)$, 则

$$Q(t+15) = \max[Q(t) + Z(t), 0] \quad (15.2)$$

于是就可利用数据估计 $Z(t)$ 的分布,同时递推 $Q(t)$ 。

3. 模拟

模型设计完成后,将进行实际的模拟,一般要经过下列过程。

(1) 确定随机变量及其分布

通常的模拟模型利用随机变量来描述真实事件,方程(15.1)和(15.2)中的 S_n 、 A_n 及 $Z(t)$ 都可视为随机变量。接下来要确定各个随机变量的分布,只有分布已知,模拟才能实现。这就需要利用统计方法计算随机变量的累积分布函数。例如设随机变量的取值区间分为 n 个间隔,且 f_i 为第 i 个间隔的频率,那么累加到第 j 个间隔并包括第 j 个间隔的累计频率

$$F_j = \sum_{i=1}^j f_i \quad (15.3)$$

(2) 产生均匀分布的随机数

要产生随机变量模拟数据,首先要获得均匀分布的随机数,即在 $[0,1]$ 区间上均匀分布的随机变量的实现值。在计算机上产生均匀随机数的常用算法是同余数法。利用均匀随机数能够产生具有其他分布随机变量的模拟数据。

(3) 产生随机变量的模拟数据

假设需要产生与问题相关的分布函数 $F(x)$ 的模拟数据,即实现随机变量 X 。这里 $F(x) = P(X \leq x)$ 。可以证明,如 R 为 $[0,1]$ 区间上均匀分布的随机变量,则有 $F(X) = R$, 因而 $X = F^{-1}(R)$, F^{-1} 表示 F 的逆函数。这意味着如果 u_n 为均匀随机数序列的第 n 个,则 $x_n = F^{-1}(u_n)$ 为 X 模拟数据的第 n 个。原则上任何分布的随机变量都可用这种变换得到,但如果逆函数难于计算,就需要用其他方法,如组合法、近似法、舍选法等。

例如,令 R 为所选的均匀随机数, x_i 为式(15.3)中随机变量取值区间中的第 i 个区间的上限,若

$$F_{i-1} < R \leq F_i$$

则该随机变量的值应取为 x_i 。

(4) 模型演算

这一步是产生在某种方式下由随机变量 X_1, X_2, \dots, X_n 所决定的响应 W , 即 $W = f(X_1, X_2, \dots, X_n)$, 例如式(15.1)和式(15.2)就说明这种情形。 W 的分布依赖于 X_i 分布的参数、问题的起始条件和模拟过程的运行长度。

4. 结果分析

利用模拟模型的结果可以研究现实系统,通过对输出数据的分析可以找出系统运行的制约因素及影响系统性状的主要变量,进而可以优化设计系统。

经过分析,如果目标已经满足,那么模拟过程也告一段落,可是因为模拟是一个反复试验不断探索的过程,所以目标常常不能满足。在这种情况下一般有两种选择:一种是修改模型,使它在模拟过程中易于辨别,然后重新进行实验;另一种是使用原来的模型,但要改变原来的试验设计重新试验。上述模拟过程如图 15.1 所示。

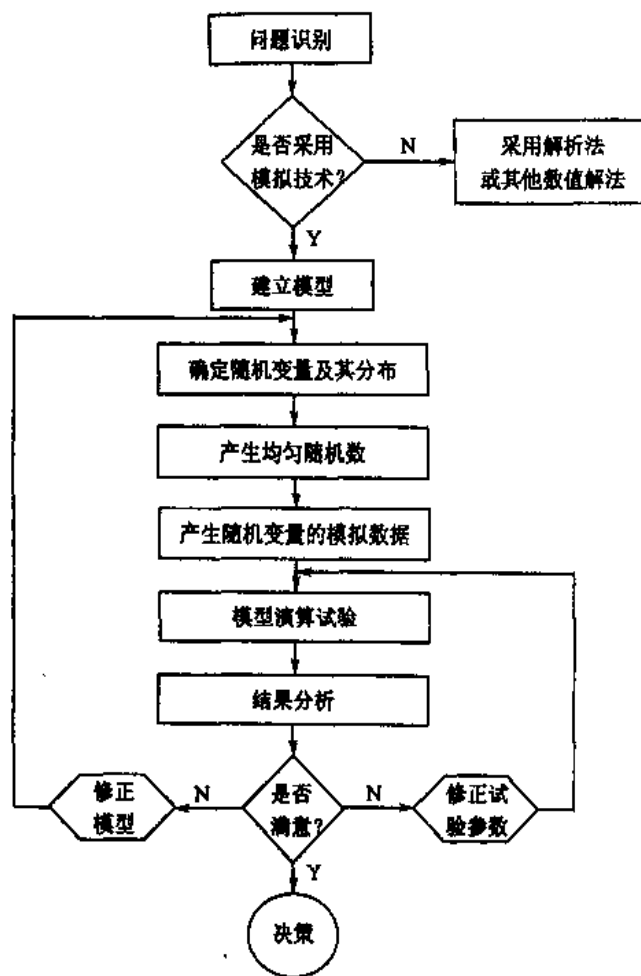


图 15.1

15.2 一个人工模拟的例子

为了实际说明模拟技术的一般原理和步骤,现举一人工模拟的例子。

1. 问题分析

某厂需外订电动机作为产品的组成部分。为保证生产的正常进行,对生产过程中所需的电动机需要有一定数量的储备,又为了使报废电动机得到补充,需要每隔一定时期提出订货。由于电动机的消耗量和订货提前期往往是随机的,因而可能发生供应上的缺货。这就需要选择一个好的存贮策略,使订货费、存贮费以及由于缺货造成的损失这三者费用总和最小。

据统计,该厂每天需要的电动机数量及频率如表 15.1 所示,订货提前期(从发出订单到收到订货的时间)及其频率如表 15.2 所示。该厂每次订货时间不定,但每次订货量均为 120 台。该厂每日对库存进行盘点,当发现库存量低于订货点储备 P 时即提出订货。已知存贮费 $K_c = 0.10$ 元/台·天,缺货损失费 $K_s = 0.50$ 元/(台·天),订购费 $R = 10$ 元/次,要求选择一个合理的订货点 P ,使该厂在订货、储备和缺货造成的损失费用的总和最少。

表 15.1

每天需要的电动机数	17	18	19	20	21	22	23
频率	0.05	0.10	0.20	0.30	0.20	0.10	0.05

表 15.2

订货提前期	1	2	3	4	5
频率	0.10	0.20	0.40	0.20	0.10

2. 建立模型

首先确定系统所涉及的变量如下:

I_k ——第 k 天(第 $k-1$ 天末)时的电动机库存量;

Q ——每批订购数量, $Q=120$;

P ——订货点储备量;

D_k ——第 k 天的需求量;

L_k ——第 k 天发出订货时的订货提前期;

C_k ——第 k 天的存贮费;

S_k ——第 k 天发生的缺货损失;

TM_k ——第 k 天开始(第 $k-1$ 天末)累计的订货费、存贮费和缺货损失费的总和;

R_k ——第 k 天发生的订货费, 即 $R_k=10$ 。

其中 $k=1, 2, \dots, n$ 为模拟天数。

根据以上分析, 得出问题的数学模型为

$$\min TM_k(P, C_k, L_k) = \sum_{k=1}^n (C_k + S_k + R_k)$$

其中 $C_k = K_C(I_k - D_k) \quad (I_k \geq D_k)$

$S_k = K_S(D_k - I_k) \quad (I_k < D_k)$

$R_k = \begin{cases} 10 & \text{第 } k \text{ 天发生订货} \\ 0 & \text{第 } k \text{ 天不发生订货} \end{cases}$

问题中的需求量 D_k 及订货提前期 L_k 是不可预测的随机因素, 用数学模型来研究是比较困难的, 而用模拟方法则能够克服这种困难。

3. 模拟

(1) 确定随机变量及其分布

本例中的需求量 D_k 及订货提前期 L_k 是随机变量。根据其频率分布来计算累积分布函数, 计算结果见表 15.3 和表 15.4, 同时进行随机数的分配。

表 15.3

每日需求量	相对频率分布	相对累积频率分布	对应的随机数
17	0.05	0.05	00~04
18	0.10	0.15	05~14
19	0.20	0.35	15~34
20	0.30	0.65	35~64
21	0.20	0.85	65~84
22	0.10	0.95	85~94
23	0.05	1.00	95~99

表 15.4

提前期	相对频率分布	相对累积频率分布	对应的随机数
1	0.10	0.10	00~09
2	0.20	0.30	10~29
3	0.40	0.70	30~69
4	0.20	0.90	70~89
5	0.10	1.00	90~99

(2)取均匀随机数

分别取均匀随机数表中的2位随机数作为 D_k 及 L_k 的随机数列,如表15.5的第3列和第5列。

(3)产生模拟数据

根据 D_k 及 L_k 的分布,产生需求量 D_k 及订货提前期 L_k 的模拟数据列如表15.5所示。

(4)模型演算试验

该例中模拟时的递推关系为

$$TM_{k+1} = TM_k + C_k + S_k + R_k$$

表15.5列出当模拟时间 $n=20$ 天、 $P=60$ 时人工模拟的过程和结果。作为方案比较,又分别计算了 $P=40、50、70$ 时的方案,结果如表15.6所示。

表 15.5

时间	I_k	每日需求量		提前期		$I_k - D_k$	C_k	S_k	R_k	TM_{k+1}
		随机数	D_k	随机数	L_k					
1	120	68	21			99	9.9			9.9
2	99	34	19			80	8.0			17.9
3	80	30	19			61	6.1			24.0
4	61	13	18			43	4.3			28.3
5	43*	70	21	53	3	22	2.2		10	40.5
6	22	55	20			2	0.2			40.7
7	122	74	21			101	10.1			50.8
8	101	30	19			82	8.2			59.0
9	82	77	21			61	6.1			65.1
10	61	40	20			41	4.1			69.2
11	41*	44	20	74	4	21	2.1		10	81.3
12	21	22	19			2	0.2			81.5
13	2	78	21			-19		9.5		91.0
14	101	84	21			80	8.0			99.0
15	80	26	19			61	6.1			105.1
16	61	04	17			44	4.4			109.5
17	44*	33	19	23	2	23	2.3		10	121.8
18	143	46	20			123	12.3			134.1
19	123	09	18			105	10.5			144.6
20	105	52	20			85	8.5			153.1

表 15.6

方案	$P=40$	$P=50$	$P=60$	$P=70$
总费用 TM	152.5	153.1	153.1	179.7

4. 结果分析

进行方案的技术经济分析,由模拟结果可知,当选择 $P=40$ 时,各项费用的总和最小。

15.3 随机变量模拟数据的产生方法

模拟中的一个关键问题是产生与真实系统相似的输入变量。这些变量大都具有随机性。因此,有效地产生具有一定分布的随机变量是一个很重要的问题。解决上述问题的基础是产生 $[0,1]$ 区间上均匀分布随机变量的样本值,即均匀随机数,其他各种分布的随机变量可以通过对均匀随机数进行适当变换而得到。

15.3.1 均匀随机数的产生方法

均匀随机数可以直接从随机数表中得到。这种方法的缺点是输入计算机的速度缓慢。产生均匀随机数的另一种方法是利用物理装置或物理过程,但这种方法的结果不可再现。在计算机模拟中使用最多的是用数学公式递推产生随机数,目前普遍采用的是同余数法。但在用手工模拟而手头又无随机数表的情况下用平方取中法较为方便。

(1)平方取中法

任取一个 $2k$ 位的数为种子,乘方后去掉头尾取中间 $2k$ 位作为第一个随机数,再取第一个随机数为种子,按相同的方法得到第二个随机数,以此类推,就可以得到一个随机数列。这种方法的缺点是容易产生退化,一旦尾数出现0后就无法清除。此外,用这种方法得到的随机数分布均匀性较差。

(2)同余数法

这种方法的递推公式为

$$x_{n+1} = (ax_n + c) \pmod{m}$$

其中 x_i 是第 i ($i = n, n+1$)个随机数; a 为乘子; c 为增量; m 为模数; x_0 为种子,它们均为非负整数; \pmod{m} 表示除以 m 所得的余数。例如当 $m=8$ 、 $a=5$ 、 $c=7$ 时,选取 x_0 为 $0\sim 7$ 之间的任意数字,就可得到如表15.7的循环的含8个数字的随机数序列。

一般选择 $m=2^b$ (b 为计算机字长), c 为奇数, $a-1$ 可被4整除, x_0 为 $0\sim m-1$ 之间的任意整数,这样得到的随机数在 $0\sim m-1$ 之间不重复。

表 15.7

x_n	x_{n+1}
4	3
3	6
6	5
5	0
0	7
7	2
2	1
1	4

15.3.2 模拟数据的产生方法

下面介绍由随机数产生其他概率分布模拟数据的一些方法。

(1)逆转换法

逆转换法基于下面的定理:设 X 是具有概率密度函数 f 及分布函数 F 的随机变量, Y 是由 $Y=F(X)$ 定义的随机变量,则 Y 在 $[0,1]$ 上服从均匀分布。

由此,要产生服从分布 F 的随机变量,就要产生 $[0,1]$ 区间上均匀分布的随机变量 R ,然后由 $F^{-1}(R)$ 求得。

例 15.1 求任意区间 $[a,b]$ 内的均匀分布随机变量的模拟数据。

解 将 $[0,1]$ 区间的均匀随机数记作 R 。 x 是 $[a,b]$ 上均匀分布的随机变量。其分布密

度函数为

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{其他} \end{cases}$$

由 $f(x)$ 可得 x 的分布函数

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

令 $R = F(x) = \frac{x-a}{b-a}$, 得 $x = a + (b-a)R$ 。

例 15.2 产生参数为 μ 的负指数分布的模拟数据。

解 负指数分布的概率密度函数

$$f(x) = \mu e^{-\mu x}$$

分布函数为

$$F(x) = 1 - e^{-\mu x}$$

令 $R = F(x) = 1 - e^{-\mu x}$, 即 $1 - R = e^{-\mu x}$, 但 $1 - R$ 也来自均匀分布, 故令

$$R = e^{-\mu x}$$

则 $x = -\frac{1}{\mu} \ln R$ 。

例 15.3 产生几何分布的模拟数据。

解 几何分布的概率密度函数为

$$f(x) = pq^x$$

式中 $p = 1 - q, x = 0, 1, 2, \dots$ 。又有

$$F(x) = \sum_{k=0}^x pq^k = (1-q) \sum_{k=0}^x q^k = 1 - q^{x+1}$$

令 $R = 1 - F(x) = q^{x+1}$, 则

$$x = \frac{\ln R}{\ln q} - 1$$

由于有些随机变量分布函数的逆函数不存在, 有些分布函数的逆函数即使存在, 但计算起来十分复杂, 因此需要用其他方法寻找模拟数据。

(2) 组合法

组合法是利用某些容易产生的随机变量来组合得到所要求的随机变量的一种方法。

例 15.4 产生泊松分布的模拟数据。

解 已知相继两个事件出现的间隔时间为负指数分布, 则在某一时间间隔内事件出现的次数服从泊松分布。根据上述关系, 可以用负指数分布的随机变量来组合产生泊松分布的模拟数据。

设序列 y_1, y_2, \dots, y_n 服从参数为 λ 的负指数分布, 则有

$$y_i = -\frac{1}{\lambda} \ln r_i$$

其中 r_i 是在 $[0,1]$ 区间上的均匀随机数。将 y_i 值按序累加,使得满足下述关系式:

$$\sum_{i=0}^x y_i \leq 1 \leq \sum_{i=0}^{x+1} y_i$$

则求得的 x 就是参数为 λ 的泊松分布的模拟数据。将 y_i 值代入上式,得

$$\sum_{i=0}^x -\frac{1}{\lambda} \ln r_i \leq 1 \leq \sum_{i=0}^{x+1} -\frac{1}{\lambda} \ln r_i$$

$$\ln \prod_{i=0}^x r_i \geq -\lambda \geq \ln \prod_{i=0}^{x+1} r_i$$

由此得 $\prod_{i=0}^x r_i \geq e^{-\lambda} \geq \prod_{i=0}^{x+1} r_i$

例 15.5 产生爱尔朗分布的模拟数据。

解 爱尔朗分布的概率密度函数为

$$f(x) = \frac{(\mu k)^k x^{k-1}}{(k-1)!} e^{-\mu k x}$$

式中 $x > 0, \mu > 0, k$ 为正整数。它是 k 个互相独立的负指数分布随机变量的和,即

$$x = \sum_{i=1}^k x_i$$

x_i 为负指数分布的随机变量。 x_i 的概率密度函数为

$$f(x_i) = \mu k e^{-\mu k x_i}$$

因 x_i 的随机序列为 $-\frac{1}{\mu k} \ln r_i$, 故有

$$x = \sum_{i=1}^k \left(-\frac{1}{\mu k} \ln r_i \right) = -\frac{1}{\mu k} \ln \prod_{i=1}^k r_i$$

由上式看出,为确定爱尔朗分布的模拟数据,需要有 k 组 $[0,1]$ 区间的均匀随机数组合而成。

例 15.6 产生二项分布的模拟数据。

解 二项分布的密度函数为

$$f(x) = C_n^x p^x (1-p)^{n-x}$$

式中 $p < 1$ 。若取 n 个 $[0,1]$ 区间的均匀随机数 r_i , 当 $r_i \leq p$ 时,令 $x_i = 1$; 当 $r_i > p$ 时,令 x_i

$= 0$, 则 $\sum_{i=1}^n x_i$ 就是二项分布的随机变量值。

(3) 近似法

这种方法一般用于随机变量的分布函数公式无法求出时的情况。

例 15.7 产生正态分布的模拟随机数列。

解 由概率论的中心极限定理知, N 个具有相同分布的独立的随机变量 R_i 的数学期望

分别为 μ_i , 方差分别为 σ_i^2 。当 N 足够大时, 量 $\sum_{i=1}^N R_i$ 趋近于期望值为 $\sum_{i=1}^N \mu_i$, 方差为 $\sum_{i=1}^N \sigma_i^2$ 的正态分布, 即

$$\sum_{i=1}^N R_i \sim N\left(\sum_{i=1}^N \mu_i, \sum_{i=1}^N \sigma_i^2\right)$$

如果令 R_i 是 $[0,1]$ 区间均匀分布的随机变量, 则

$$f(r_i) = 1 \quad (0 \leq r_i \leq 1)$$

$$\mu_i = E(R_i) = \int_0^1 x \cdot 1 \cdot dx = \frac{1}{2}$$

$$\sigma_i^2 = \int_0^1 [x - E(x)]^2 \cdot 1 \cdot dx = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{12}$$

当 N 足够大时有

$$\sum_{i=1}^N R_i \sim N\left(\sum_{i=1}^N \frac{1}{2}, \sum_{i=1}^N \frac{1}{12}\right) = N\left(\frac{N}{2}, \frac{N}{12}\right)$$

为了得到标准的正态分布, 令

$$Z = \frac{\sum_{i=1}^N R_i - \frac{N}{2}}{\sqrt{N/12}} \sim N(0, 1)$$

在实际应用时, 当 $N=12$ 已足够精确了。因此, 用 $N=12$ 代入上式得

$$Z = \left(\sum_{i=1}^{12} R_i - 6\right) \sim N(0, 1)$$

因此只要分别给出 12 个相互独立的 $[0, 1]$ 区间的均匀随机数序列, 就可以得到一个标准的正态分布模拟数据。

若 X 为期望值 μ 、方差 σ^2 的正态分布, 因有

$$\frac{X - \mu}{\sigma} = Z$$

故

$$X = \mu + \sigma Z = \mu + \sigma \left(\sum_{i=1}^{12} R_i - 6\right)$$

(4) 舍选法

当随机变量分布函数的逆函数不存在, 又不能用组合法、近似法时, 通常采用舍选法。令 $f(x)$ 为概率密度函数, 且 $0 \leq f(x) \leq M, a \leq x \leq b$, 则舍选法的步骤为:

- ① 产生两个均匀随机变量 R_1, R_2 ;
- ② 建立 $f(x)$ 的预随机变量 $x = a + (b - a)R_1$;
- ③ 检验 $R_2 \leq f[a + (b - a)R_1]/M$ 是否成立;
- ④ 若不等式成立, 则 $x = a + (b - a)R_1$ 作为模拟数据;
- ⑤ 若不等式不成立, 则产生两个新的随机数重新进行检验。

例 15.8 产生 β 分布的随机模拟数据。

解 β 分布由下面的密度函数定义:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (0 \leq x \leq 1)$$

其中 α 和 β 是分布参数, 并且

$$\Gamma(\phi) = \int_0^\infty x^{\phi-1} e^{-x} dx$$

首先求出:

$$M = \frac{\alpha - 1}{\alpha + \beta - 2}$$

仅当 α 和 β 有确定值时, M 才有上述关系。具体步骤如下:

- ①选择 α 和 β ;
- ②计算 $M = (\alpha - 1) / (\alpha + \beta - 2)$;
- ③产生两个随机数 R_1 和 R_2 ;
- ④若 $R_2 M \leq f(x \equiv R_1) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} R_1^{\alpha-1} (1 - R_1)^{\beta-1}$ 成立, 则 R_1 作为 β 分布的模拟数据;
- ⑤若 $R_2 M > f(x \equiv R_1)$, 则回到第④步重新开始。

由于舍选法只需知道概率密度函数就可以产生反映这种概率分布的模拟数据, 因此适用性很广。但它的缺点是, 由于对每个试验点都必须计算两个均匀分布随机数, 加上有些点被舍去, 因此产生一个数据平均需要计算两个以上的均匀随机数, 加大了工作量; 另外, 应用舍选法时要求概率密度函数的变量有上下限, 而一般的概率密度函数大多具有很长的尾数, 并常常扩展到无穷大, 因此需要截断。截断后将引起误差。但当误差很小时, 可以忽略不计。

15.4 模拟语言简介

模拟语言是一种工具。正确使用模拟语言, 是模拟设计成败的关键环节, 它可以保证模拟研究的经济可行性, 节省大量人力。

因为模拟模型不遵循任何格式化的结构, 所以在模拟程序语言中使用各种各样的方法。现简单介绍几种常用的模拟语言。

GPSS 是 G. Cordon 在 60 年代初期为 IBM 公司设计的模拟语言, 它要求编译人员用模块指令建立系统的逻辑模型。模块指令将实现这种语言特有的功能。GPSS 比较适合研究存贮和排队问题。

SIMSCRIPT 是 60 年代初期由兰德公司的 H. M. Markowitz 提出, 仅适用于特定的计算机系统。SIMSCRIPT 所描述的系统仅在一定的事件时间上才发生状态变化, 这个系统用状态变化事件描述, 并由实体组成。SIMSCRIPT 适用于事件间隔模拟模型。

GASP II 是 60 年代中期由 Pritsker 在亚利桑那州大学开发的。GASP II 把系统描述为包含各种状态变化的实体。这些实体叫做要素。每个要素都具有属性。事件的发生取决于系统状态、属性值或属性的集合。GASP II 具有很强的探测错误系统和灵活的输出格式。

SIMULA 是 O. Dall 和 K. Nyggard 为兰德公司设计的。它的一个特点是能够产生、取消和修正当前过程和由大量 SIMULA 控制语句建立的新过程。SIMULA 提供固定格式的输出和错误检测装置。

DYNAMO 是 60 年代初期由麻省理工学院的 R. Fox 和 A. L. Pugh 提出, 专门用于研究某些由一组有限差分方程描述的系统。DYNAMO 的实质是由固定时间推进机制把固有的连续函数关系和运行特性离散化。DYNAMO 至今已有效地用于经济模型、大工业中心的模拟及都市、社会、世界系统模型。

GASPIV 是由 Pritsker 和 N. Hurst 于 1973 年建立。它可以编制离散的、连续的或两者均有的模拟程序。GASPIV 把 GASP II 的全部概念具体化了, 两者的结构形式相同, 而主要区别是对状态空间事件的定义。GASPIV 已成为广泛应用和普遍接受的模拟语言。

除前面介绍的模拟语言外, 还有其他多种语言。它们或有专门用途, 或在小范围内使用。

习 题

15.1 某厂生产一种新产品,投资为 5 000 元,现需预测未来一年的销售获利情况。有三个不确定因素,即售价、可变成本、年销售量。但已掌握这三个(相互独立的)因素的不同水平的概率分布,如表 15.8、15.9、15.10。

表 15.8

售价(元)	概率
4	0.3
5	0.5
6	0.2

表 15.9

单位可变成本(元)	概率
2	0.1
3	0.6
4	0.3

表 15.10

年销售量(件)	概率
3 000	0.2
4 000	0.4
5 000	0.4

现用随机模拟的方法做预测,模拟 25 次试验,由随机数表取三组随机数 RN_1 、 RN_2 、 RN_3 依次模拟售价、单位成本和销售量,如表 15.11 所示。

表 15.11

试验号码	RN_1	售价(元)	RN_2	成本(元)	RN_3	销售量(件)	纯利润(千元)
1	8		0		6		
2	0		4		3		
3	6		2		2		
4	1		4		0		
5	3		6		0		
6	5		6		9		
7	1		6		7		
8	3		8		6		
9	2		8		8		
10	1		6		1		
11	5		7		3		
12	9		9		6		
13	4		9		7		
14	7		2		6		
15	9		5		3		
16	0		5		0		
17	1		1		8		
18	0		6		4		
19	8		8		6		
20	9		2		4		
21	0		7		7		
22	0		0		8		
23	4		0		1		
24	6		5		8		
25	4		0		1		

(1)试依照公式

$$\text{纯利润} = (\text{售价} - \text{单位可变成本}) \times \text{销售量} - 5\,000$$

计算各次试验模拟的纯利润和平均纯利润。

(2)分别计算下列三种情况发生的可能性(步骤、百分数):①遭受损失;②不受损失但纯利润不到 5 000 元;③纯利润在 5 000 元及其以上。

15.2 假设一火车修理厂考虑改造油漆分厂。新厂对火车机头和车厢外部油漆采用自动化作业线,按工序分为除漆和喷漆两部分。无论车头还是车厢均先在第 I 区(除漆)前排队等待其空闲,而在除漆后在原工区

内等待第Ⅱ区(喷漆)空闲再进行喷漆。车头和车厢所占百分比以及在不同区内所需时间如表 15.12。

表 15.12

类型	比例(%)	需要时间(h)	
		I 区	Ⅱ区
车头	20	2	3
车厢	80	1	2

车头和车厢到达油漆分厂的数量服从泊松分布,平均每 4 小时有一辆到达,其中 20% 是车头,其余为车厢,两类是随机混合的。

由于在Ⅱ区需要工时较多,提出两种不同方案:

A_1 在每区内设置 1 辆车的加工位置。

A_2 在Ⅰ区设置 1 辆车的加工位置,在Ⅱ区设置 2 辆车的加工位置。

(1)利用表 15.13 和表 15.14 给出的表格模拟油漆分厂 30 个工时内到达车头和车厢的工作情况(在两种方案下)。(注:①保证在 30 h 内到达的车辆完成油漆工作,即 30 h 后不允许再进车辆。②工时是用百分小数表示的。)

表 15.13 方案 1(A_1)

到达		类型		进入时间		离去时间	系统内停留时间
随机数	时间(h)	随机数	类型	I 区	Ⅱ区		
43	2.25	75					
29	3.62	72					
11	4.09	04					
50	6.86	75					
98	22.51	41					
14	23.11	08					
24	24.21	25					
61	27.98	24					
54	31.08	36					
02	31.16	27					
35	32.88	92					

表 15.14 方案 2(A_2)

到达		类型		进入时间		离去时间	系统内停留时间
随机数	时间(h)	随机数	类型	I 区	Ⅱ区		
43	2.25	75					
29	3.62	72					
11	4.09	04					
50	6.86	75					
98	22.51	41					
14	23.11	08					
24	24.21	25					
61	27.98	24					
54	31.08	36					
02	31.16	27					
35	32.88	92					

(2)计算在两种方案下,两类车在油漆分厂内的平均时间。

(3)假设方案 2 比方案 1 固定资产投资多 500 000 元,而车头和车厢空闲 1 h 费用分别为 1 000 元和 400 元,油漆分厂每年有 8 760 个工时。如果方案 2 在一年内可收回固定资产多投资部分,则管理人员认为可以采纳。计算每年各类车辆来油漆分厂喷漆的数量。请问你建议采用何种方案,说明理由。

参 考 文 献

- 1 钱颂迪. 运筹学(修订版). 北京:清华大学出版社,1990
- 2 胡运权. 运筹学教程. 北京:清华大学出版社,1998
- 3 胡运权. 运筹学习题集. 北京:清华大学出版社,1995
- 4 席少霖. 非线性最优化方法. 北京:高等教育出版社,1996
- 5 袁亚湘. 非线性数值方法. 上海:上海科学技术出版社,1993
- 6 魏权龄. 数学规划引论. 北京:航空航天大学出版社,1991
- 7 E Turban, J R Meredith. Fundamentals of Management Science. 5Th ed. R. P. Donnelley & Sons Company, USA, 1991
- 8 F S Hillier, G J Lieberman. Introduction to Operations Research. 4Th ed. Holden-Day, Inc, USA, 1986
- 9 盛昭瀚等. DEA 理论、方法与应用. 北京:科学出版社,1996
- 10 吴育华等. 管理数学. 天津:天津科技翻译出版公司,1991
- 11 于九如等. 应用统计学. 天津:天津科技翻译出版公司,1991
- 12 胡毓达. 实用多目标最优化. 上海:上海科学技术出版社,1990
- 13 刘德铭等. 对策论及其应用. 长沙:国防科技大学出版社,1995
- 14 范贻昌. 实用管理运筹学. 天津:天津大学出版社,1995
- 15 吴育华等. 工程经济数学方法. 天津:天津大学出版社,1989
- 16 顾基发,魏权龄. 多目标决策问题. 应用数学与计算数学,1980,(1)
- 17 王浣尘. 可行性研究和多目标决策. 北京:机械工业出版社,1986
- 18 许树柏. 层次分析法原理. 天津:天津大学出版社,1988
- 19 林铿云,董加礼. 多目标决策的方法与理论. 长春:吉林教育出版社,1992
- 20 韩崇昭,张平平. 决策、对策与管理. 北京:新时代出版社,1986
- 21 王莲芬,许树柏. 层次分析法引论. 北京:中国人民大学出版社,1990
- 22 T L 萨蒂著. 层次分析法在资源分配、管理、冲突分析中应用. 北京:北京煤炭工业出版社,1990
- 23 魏权龄. 评价相对有效性的 DEA 方法. 北京:中国人民大学出版社,1988
- 24 吴育华,李丛东. DEA 方法与生产函数. 系统工程,1995,13(3)
- 25 陈珽. 决策分析. 北京:科学出版社,1989
- 26 王建华. 对策论. 北京:清华大学出版社,1986
- 27 王伟勋. 管理对策分析. 上海:上海交通大学出版社,1987
- 28 张维迎. 博弈论与信息经济学. 上海:上海人民出版社,1996
- 29 张守一. 现代经济对策论. 北京:高等教育出版社,1998
- 30 Charnes A, Cooper W W. Cone Ratio Data Envelopment Analysis and Multi-objective Programming. International Journal of Systems Science, 1989,20(7)

- 31 Y Sawaragi, H Nakayama, T Tanino. Theory of Multi-objective Optimization. Mathematics in Science and Engineering, 1985,176
- 32 Ralph Steuer E. Multiple Criteria Optimization Theory. Computation and Application. John wiley & son, 1996
- 33 S H Tijs and T S H Driessen. Game Theory and Cost Allocation Problems. Management Science, 1986,32(8)